MESH-HR: Multimodal Fusion of Histopathology and Structured Somatic Genomics for Continuous Breast Cancer Receptor Subtyping

Shaye Carver [1 2] **Kodi Taraszka** [1] **Intae Moon** [1] **Alexander Gusev** [1]

Abstract

Breast cancer receptor subtyping determines eligibility for targeted therapies based on estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) expression, which are typically assessed via immunohistochemistry (IHC) for ER and PR, and IHC or fluorescence in situ hybridization (FISH) for HER2. We introduce MESH-HR (Multimodal Ensemble of Somatic Variants and H&E Slides for Hormone Receptor subtyping), a multimodal framework that integrates two biological modalities, somatic DNA profiles and whole-slide histopathology images, to predict receptor status probabilities without molecular marker assays (IHC/FISH). To enable large-cohort training without manual curation, ground truth receptor labels are extracted at scale from unstructured pathology reports using a GPT-4–based LLM pipeline. MESH-HR combines an attention-based multiple instance learning (ABMIL) vision encoder with an XGBoost model over somatic features, fusing complementary signal across DNA and imaging modalities. Trained on $>$1,300 breast cancers, MESH-HR achieves AUCs of 0.94 (ER), 0.83 (PR), and 0.96 (HER2) on held-out data, outperforming unimodal baselines, and generalizes zero-shot to The Cancer Genome Atlas Breast Cancer cohort (TCGA-BRCA). Continuous probabilistic outputs improve survival stratification over binary clinical labels, recovering prognostic signal lost under discrete IHC thresholds. We further apply MESH-HR zero-shot to cancers of unknown primary (CUP), a population in which receptor profiling is rarely performed due to the absence of a known tissue of origin, obtaining biologically consistent, survival-predictive receptor estimates that support receptor-informed personalized treatment stratification.

[1]Dana-Farber Cancer Institute, Boston, MA, USA [2]Harvard Medical School, Boston, MA, USA. Correspondence to: Shaye Carver <scarver@g.harvard.edu>.

Proceedings of the 43rd International Conference on Machine Learning, Seoul, South Korea. PMLR 306, 2026. Copyright 2026 by the author(s).

1. Introduction

Breast cancer subtyping is defined by the expression status of three biomarkers — ER, PR, and HER2 — which determine eligibility for endocrine therapy, CDK4/6 inhibitors, and anti-HER2 agents, and are assessed via molecular marker assays (IHC for ER/PR; IHC/FISH for HER2) ordered only when a tissue of origin is already known. Cancer of unknown primary (CUP) accounts for 3–5% of all malignancies and carries poor prognosis under empiric chemotherapy (1). A subset of CUP patients harbor receptor-expressing tumors spanning breast and other hormone- or HER2-associated lineages, yet without a known primary site, standard receptor profiling is rarely ordered. These patients cannot be identified, and effective targeted therapies go unprescribed.

Beyond CUP, standard molecular marker assays impose a second limitation: clinical practice binarizes receptor expression as positive or negative, discarding prognostically meaningful variation along a biological continuum (2; 3; 4). Borderline cases are poorly served by hard thresholds, and variability in assays and interpretation introduces noise into the label itself. Continuous probabilistic outputs offer a more faithful representation of this spectrum, enabling uncertainty-aware risk stratification that binary calls cannot provide.

Related work. Prior work has demonstrated that hematoxylin and eosin (H&E) morphology can predict molecular subtypes in breast cancer (5; 6), and somatic mutation profiles have been used for tissue-of-origin classification (1). However, no existing approach jointly fuses both modalities to infer receptor status, outputs calibrated continuous probabilities, or is validated in IHC-naive CUP cohorts. EXPAND (7), the most relevant imaging-only baseline, classifies H&E whole-slide images (WSIs) into HR+, HER2+, and triple negative breast cancer (TNBC) subtypes using nuclear morphological features, but does not incorporate structured genomic data or full-slide context.

This work. We introduce MESH-HR to address both limitations by estimating continuous receptor subtype probabilities from data available in CUP: H&E WSIs and structured somatic genomic tables. MESH-HR frames receptor subtyping as multimodal learning over high-dimensional somatic features and multi-resolution image signals, using a two-branch ensemble that pairs an ABMIL vision encoder with an XGBoost genomic model and performs receptor-specific fusion. Across ER, PR, and HER2, MESH-HR outperforms unimodal models on internal and external validation. We further show that its continuous scores stratify survival in known-primary breast cancers better than binarized predictions or clinical labels. In CUP, where IHC labels are often unavailable, MESH-HR predictions are concordant with independently inferred tissue lineage and stratify overall survival. Finally, discordance analyses show that when MESH-HR and clinical labels disagree, patient outcomes align more closely with the model's predictions than with the recorded labels.

2. Method

Data and labels. We assembled 1,368 breast cancer cases (cancers of known primary; CKPs) from the Dana-Farber Cancer Institute PROFILE cohort, each with a paired H&E WSI and somatic genomic profile. Clinical receptor labels (ER, PR, HER2) were extracted from unstructured free-text pathology reports using a GPT-4-based LLM extraction pipeline; a random subset was manually reviewed to verify accuracy. Zero-shot generalization was evaluated on 786 CUP cases from the same cohort, for which no IHC receptor labels were available.

Vision branch. H&E WSIs were segmented and tiled into 256×256 patches using TRIDENT (8; 9); the vision branch takes as input patch-level embeddings generated by UNI (10), aggregated into a slide-level representation via a gated Attention-Based Multiple Instance Learning (ABMIL) model (11; 12).

Genomics branch. The genomics branch takes as input a high-dimensional tabular feature matrix comprising binary somatic mutation indicators per gene, gene-level copy number alteration (CNA) scores, and mutational signature weights, and fits an XGBoost classifier (13) per receptor.

Fusion. Branch predictions are combined via a receptor-specific ensemble selected by out-of-sample cross-validation performance: a weighted average for ER and HER2, and a logistic meta-learner for PR, chosen from candidates including logistic-regression stacking and optimized weighted averages. MESH-HR outputs both a continuous probability and a thresholded binary call per receptor.

Table 1. AUC for per-receptor classification. Internal: held-out test set. External: zero-shot evaluation on TCGA-BRCA without retraining.

Model	Cohort	ER	PR	HER2
ABMIL (H&E only)	Internal	0.93	0.78	0.80
XGBoost (genomics)	Internal	0.88	0.80	0.95
MESH-HR	Internal	**0.94**	**0.83**	**0.96**
MESH-HR	TCGA-BRCA	0.90	0.91	0.94

Training and evaluation. Independent models were trained per receptor and tuned via randomized hyperparameter search within five-fold nested cross-validation; final performance was evaluated on a held-out test set. For analyses spanning the full breast cohort, out-of-sample predictions were extracted via cross-validation to avoid in-sample evaluation while retaining maximal sample size.

3. Experiments

3.1. Receptor-Specific Multimodal Fusion Outperforms Unimodal Approaches

MESH-HR consistently achieves the highest AUC across all three receptors on the internal held-out test set (ER 0.94, PR 0.83, HER2 0.96), outperforming both unimodal baselines (Table 1). No single modality was uniformly optimal across receptors: ER prediction was driven primarily by morphologic features (ABMIL AUC 0.93), PR benefited most from multimodal integration, and HER2 was captured primarily by somatic genomic alterations (XGBoost AUC 0.95). Receptor-specific fusion allowed MESH-HR to adaptively weight histologic and genomic contributions rather than enforcing a fixed ensemble, recovering the strongest predictive signal in each setting. The structured somatic genomic branch provides complementary signal beyond morphology, most notably for ER and HER2. Interpretability analysis confirms biologically grounded learning: in the vision branch, high-attention patches are enriched for tumor tissue over stroma; in the genomics branch, HER2 prediction is dominated by *ERBB2* CNA, ER prediction by *TP53* mutation, and *ESR1* (the gene encoding ER) ranks low in feature importance, consistent with its role in metastatic rather than primary disease.

Applied without retraining to TCGA-BRCA (14), MESH-HR achieves zero-shot AUCs of 0.90 (ER), 0.91 (PR), and 0.94 (HER2), largely on par with internal performance. TCGA-BRCA cases were subsampled to match internal class distributions. For TCGA-BRCA evaluation, a ResNet50-based vision encoder was used in place of UNI, as its lower-capacity representations generalize more robustly across the heterogeneous image quality present in multi-institutional cohorts. These results exceed published

Table 2. Subtype classification: MESH-HR vs. EXPAND on internal cohort. Balanced accuracy and macro F1 are primary metrics given binary predictions only.

Phenotype	Model	Bal. Acc	Macro F1
HER2+	MESH-HR	**0.91**	**0.93**
HER2+	EXPAND	0.50	0.48
HR+	MESH-HR	**0.81**	**0.83**
HR+	EXPAND	0.50	0.23
TNBC	MESH-HR	**0.65**	**0.69**
TNBC	EXPAND	0.50	0.19

imaging-only methods evaluated on TCGA: ReceptorNet (5) achieves ER AUC 0.85 (zero-shot) and PR AUC 0.81 (partially trained on TCGA); Farahmand et al. (15) achieve HER2 AUC 0.81 (zero-shot). MESH-HR exceeds all three baselines despite training on an independent cohort, predicting all receptors simultaneously, and incorporating structured genomic data.

We further compare against EXPAND (7), a published imaging-only method that classifies H&E WSIs into HR+ (hormone receptor positive; ER or PR positive), HER2+, and TNBC (triple negative breast cancer; ER, PR, and HER2 negative) using nucleus-level features. To match this setting, MESH-HR predictions are thresholded at 0.5 and mapped to the same phenotype definitions. Because evaluation is performed on discrete multi-class labels rather than continuous scores, we report balanced accuracy and macro F1 (Table 2).

EXPAND achieves balanced accuracy of 0.50 across all shared subtypes, equivalent to random chance, while MESH-HR achieves 0.91, 0.81, and 0.65 for HER2+, HR+, and TNBC respectively, with macro F1 scores of 0.93, 0.83, and 0.69. The near-random EXPAND performance likely reflects domain shift between the cohort on which EXPAND's nuclear features were developed and the internal cohort slides. More broadly, EXPAND's reliance on explicitly defined nuclear features limits the predictive signal to a single cellular compartment, whereas MESH-HR captures tissue architecture, stromal context, and spatial organization via ABMIL, further extended by structured genomic data.

3.2. MESH-HR Reveals Prognostic Information Beyond Clinical Receptor Status

Beyond classification, MESH-HR's multimodal predictions act as digital biomarkers of clinical outcome. Breast cancer survival is known to differ by hormone receptor status, both because ER/PR+ tumors are biologically less aggressive (16) and because ER/PR+ and HER2+ tumors are amenable to targeted therapies. We asked whether MESH-HR predictions are more prognostic than the clinical labels used as training ground truth, which carry inherent noise from variability in IHC staining, pathologist interpretation, and documentation. We generated out-of-fold MESH-HR predictions for each breast CKP using 5-fold cross-validation and fit multivariable Cox proportional hazards models adjusted for age, tumor purity, biopsy site and type, and sequencing panel version.

Comparing three receptor representations—binary clinical labels, binary MESH-HR predictions, and continuous MESH-HR probabilities—continuous MESH-HR predictions provided the greatest discrimination of overall survival (C-index $= 0.661$), followed by binary MESH-HR predictions (C-index $= 0.652$) and binary clinical labels (C-index $= 0.646$). Adding binary clinical labels to continuous MESH-HR predictions yielded only marginal improvement (C-index $= 0.667$), indicating that continuous predictions largely subsume the prognostic information in clinical labels. Individual receptor associations were directionally consistent across representations, with receptor positivity protective in all models (all hazard ratios < 1.0). These results indicate that MESH-HR predictions stratify breast cancer survival comparably to or better than clinical receptor labels, and that binary thresholding discards meaningful prognostic signal that continuous probabilistic predictions preserve.

3.3. MESH-HR Generalizes Zero-Shot to Cancer of Unknown Primary

To evaluate out-of-distribution generalization in a clinically realistic, label-sparse setting, we applied MESH-HR to CUP without retraining. CUP tumors lack known tissue of origin and often lack receptor IHC measurements; to contextualize predictions, we used an independent tissue-of-origin model developed in our laboratory (manuscript in preparation) to infer tumor lineage. Among tumors inferred to arise from breast lineages ($n = 27$), a covariate-adjusted Cox model showed that MESH-HR-predicted HR-positive status (defined as ER- or PR-positive) was strongly associated with improved overall survival (hazard ratio $= 0.092$, $p = 0.003$). Given that HR-positive breast cancers are typically associated with more favorable survival than HR-negative tumors, this finding is consistent with established breast cancer biology and demonstrates that MESH-HR captures biologically meaningful prognostic information in CUP tumors inferred to arise from breast lineages.

Across all CUP tumors ($n = 660$), tumors were stratified by their model-inferred tissue lineage of origin into HR-expressing and non-HR-expressing tissues, spanning a broad range of predicted primary sites. MESH-HR predictions exhibited strong lineage specificity: ER and PR probabilities were significantly higher in HR-expressing lineages (ER $p = 1.60 \times 10^{-9}$; PR $p = 1.26 \times 10^{-12}$; one-tailed *t*-test), and HER2 probabilities were elevated in

HER2-expressing lineages ($p = 5.01 \times 10^{-9}$), indicating that MESH-HR preferentially assigns receptor signals to biologically appropriate lineage contexts. Consistent with this lineage specificity, a Cox model incorporating an interaction term between predicted HR status and inferred HR-lineage (HR-expressing vs. non-HR-expressing lineages) confirmed lineage-dependent prognostic effects (interaction hazard ratio $= 0.503$, $p = 0.006$): HR-positive predictions were associated with improved survival in HR-expressing lineages but not in non-HR-expressing ones (Figure 1A). Notably, MESH-HR was trained exclusively on breast cancer specimens, yet receptor probabilities were preferentially elevated in CUPs assigned to receptor-expressing lineages and were prognostic only within those lineages, suggesting the model captures histologic features of hormone receptor biology that generalize beyond breast-specific morphology. Together, these results show that MESH-HR generalizes zero-shot to CUP and recovers lineage-specific, clinically meaningful survival structure, enabling receptor-informed treatment stratification in a setting where such profiling is rarely performed.

3.4. Discordant Clinical and MESH-HR Classifications Follow MESH-HR-Predicted Outcomes

To assess whether MESH-HR captures tumor biology beyond clinical labels, we stratified breast CKPs by agreement between MESH-HR-predicted HR positive status and clinical HR status into four concordance groups: concordant (MESH-HR+/label+, MESH-HR−/label−) and discordant (MESH-HR−/label+, MESH-HR+/label−), where HR+ is defined as ER- or PR-positive. HR+ tumors are typically associated with more favorable survival than HR− tumors. Consistent with established breast cancer biology, concordant HR+ patients exhibited the most favorable outcomes in covariate-adjusted Cox models, whereas concordant HR− tumors had worse survival (hazard ratio $= 2.31$, $p = 2.0 \times 10^{-16}$). Notably, survival among discordant cases aligned more closely with MESH-HR–derived status than with the clinical classification. Discordant MESH-HR+/label− cases showed survival comparable to the concordant HR+ group and were not significantly different (hazard ratio $= 1.35$, $p = 0.057$), whereas discordant MESH-HR−/label+ cases experienced significantly worse survival (hazard ratio $= 1.81$, $p = 0.007$), approaching that of HR− disease (Figure 1B).

We further restricted analysis to clinically HR^+ patients receiving endocrine therapy. Among these, discordant MESH-HR^-/$label^+$ patients experienced significantly worse survival following endocrine therapy initiation than concordant HR^+ patients (hazard ratio $= 2.30$, $p = 0.025$), suggesting that MESH-HR identifies a subset of clinically HR^+ tumors whose biology more closely resembles HR^- disease. Notably, nearly all discordant MESH-HR−/label+ tumors were grade 3, indicating that image-based features associated with aggressive tumor biology drive these reclassifications. Given the well-established association between high histologic grade and hormone receptor-negative breast cancer, the predominance of grade 3 tumors among discordant MESH-HR^-/$label^+$ cases suggests that MESH-HR is capturing underlying tumor biology that is more consistent with HR^- disease than with their clinical receptor labels. Collectively, these analyses indicate that when MESH-HR disagrees with the clinical label, patient outcomes follow the model's prediction, supporting the ability of MESH-HR to

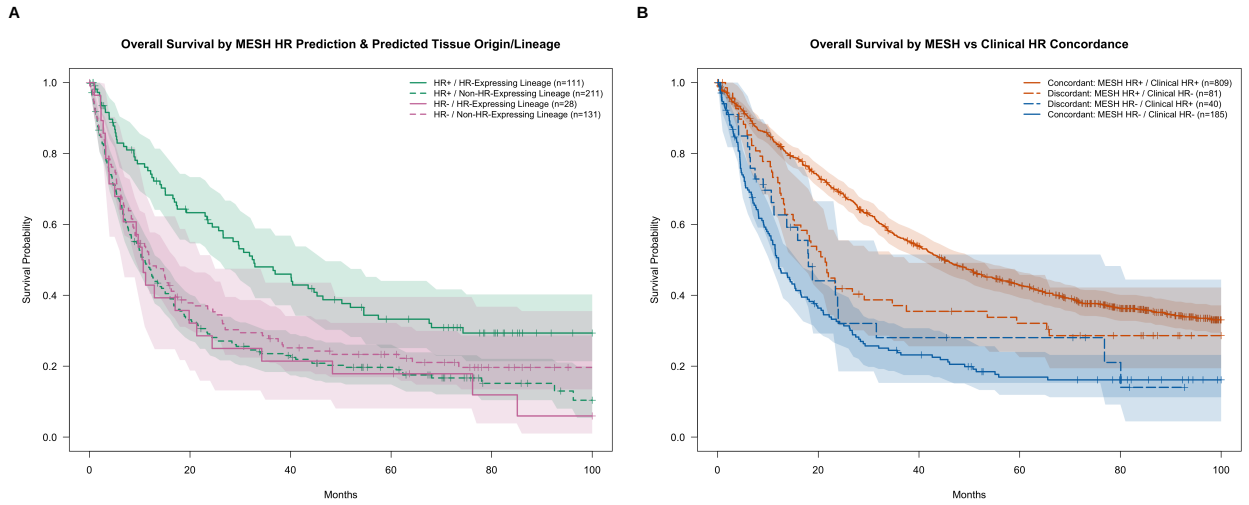

Figure 1. Overall survival stratified by MESH-HR prediction. Kaplan–Meier curves for (A) CUP patients stratified by predicted HR status and model-inferred tumor lineage (HR-expressing vs. non-HR-expressing), and (B) breast CKPs stratified by concordance between clinical HR labels and MESH-HR predictions.

capture clinically meaningful variation in hormone receptor biology not fully reflected by routine clinical testing.

4. Conclusion

We present MESH-HR, a multimodal framework trained on known-primary breast cancers to infer continuous ER, PR, and HER2 receptor probabilities from H&E whole-slide images and structured somatic genomic profiles without requiring receptor IHC at inference. Through receptor-specific fusion, MESH-HR combines complementary morphologic and genomic signals, outperforming unimodal baselines and generalizing zero-shot to TCGA-BRCA, demonstrating robust receptor prediction across breast cancer cohorts.

MESH-HR predictions also capture clinically meaningful biology beyond classification. Continuous receptor probabilities improved survival stratification relative to binary clinical labels, and in discordant cases, patient outcomes followed MESH-HR-predicted status rather than the recorded clinical label. When applied zero-shot to CUP, where tissue of origin and receptor labels are often unavailable, MESH-HR predictions were enriched in biologically appropriate model-inferred lineages and were prognostic only within receptor-expressing lineage contexts. Together, these findings support MESH-HR as a scalable approach for receptor-informed risk stratification in known-primary breast cancer and label-sparse CUP. Future work should prioritize prospective validation, probability calibration, and evaluation of treatment-response endpoints before clinical deployment.

References

[1] Intae Moon et al. Machine learning for genetics-based classification and treatment response prediction in cancer of unknown primary. *Nature Medicine*, 29:2057–2067, 2023.

[2] Shanu Modi et al. Trastuzumab deruxtecan in previously treated HER2-low advanced breast cancer. *New England Journal of Medicine*, 387:9–20, 2022.

[3] Ming Yang et al. Clinicopathological characteristics of HER2-low breast cancer: a retrospective study. *Scientific Reports*, 13:12382, 2023.

[4] Rachel Yoder et al. Impact of low versus negative estrogen/progesterone receptor status on clinicopathologic characteristics and survival outcomes in HER2-negative breast cancer. *npj Breast Cancer*, 8:80, 2022.

[5] Nikhil Naik et al. Deep learning-enabled breast cancer hormonal receptor status determination from base-level H&E stains. *Nature Communications*, 11:5727, 2020.

[6] Gil Shamai 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*, 4:276, 2024.

[7] Rahul Kumar Barman et al. Pathologist-interpretable breast cancer subtyping and stratification from AI-inferred nuclear features. *bioRxiv*, 2025.

[8] Andrew Zhang, Guillaume Jaume, Anurag Vaidya, Tong Ding, and Faisal Mahmood. Accelerating data processing and benchmarking of ai models for pathology. *arXiv preprint arXiv:2502.06750*, 2025.

[9] Anurag Vaidya, Andrew Zhang, Guillaume Jaume, Andrew H Song, Tong Ding, Sophia J Wagner, Ming Y Lu, Paul Doucet, Harry Robertson, Cristina Almagro-Perez, et al. Molecular-driven foundation model for oncologic pathology. *arXiv preprint arXiv:2501.16652*, 2025.

[10] Richard J Chen, Tong Ding, Ming Y Lu, Drew F K Williamson, Akhil Sharma, Bowen Chen, Daniel Shao, Muhammad Shaban, Mane Williams, Babak Ehteshami Bejnordi, et al. Towards a general-purpose foundation model for computational pathology. *Nature Medicine*, 30:850–862, 2024.

[11] Hao Hu, Rui Ye, Jeyan Thiyagalingam, Frans Coenen, and Jionglong Su. Triple-kernel gated attention-based multiple instance learning with contrastive learning for medical image analysis. *Applied Intelligence*, pages 1–16, 2023.

[12] Maximilian Ilse, Jakub M. Tomczak, and Max Welling. Attention-based deep multiple instance learning. *arXiv preprint arXiv:1802.04712*, 2018.

[13] Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016.

[14] The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*, 490:61–70, 2012.

[15] Saman Farahmand, Aileen I. Fernandez, Fahad Shabbir Ahmed, David L. Rimm, Jeffrey H. Chuang, Emily Reisenbichler, and Kourosh Zarringhalam. Deep learning trained on hematoxylin and eosin tumor region of interest predicts HER2 status and trastuzumab treatment response in HER2+ breast cancer. *Modern Pathology*, 35:44–51, 2021.

[16] Harold J. Burstein. Systemic therapy for estrogen receptor-positive, HER2-negative breast cancer. *New England Journal of Medicine*, 383:2557–2570, 2020.