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

Anonymous Authors¹

Abstract

Breast cancer receptor subtyping guides treatment and prognosis, as eligibility for targeted therapies is determined by estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) expression or amplification status. While these biomarkers are clinically assessed using immunohistochemistry (IHC), computational approaches have largely focused on directly predicting receptor status from histopathology images, despite complementary signal in somatic genomic profiles. We introduce MESH-HR (Multimodal Ensemble of Somatic Variants and H&E Slides for Hormone Receptor subtyping), a model that integrates structured somatic genomic features and multi-resolution pathology images to predict ER, PR, and HER2 receptor status probabilities from routinely collected data, without requiring IHC. MESH-HR combines an attention-based multiple instance learning (ABMIL) vision encoder with an XGBoost model over structured genomics, capturing complementary signal across modalities. Trained on >1,300 breast cancers, MESH-HR achieves AUCs of 0.90 (ER), 0.84 (PR), and 0.96 (HER2) on held-out data, outperforming unimodal and prior imaging-only approaches, and generalizes zero-shot to The Cancer Genome Atlas Breast Cancer cohort (TCGA-BRCA). Continuous predictions improve survival stratification over binary outputs and discretized clinical IHC categories, recovering signal lost under discrete clinical labeling. We further apply MESH-HR to cancers of unknown primary (CUP), where receptor status is often unavailable, obtaining biologically consistent, survival-predictive subtype estimates that enable receptor-informed stratification.

¹Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

1. Introduction

Breast cancer treatment is guided by three key biomarkers (ER, PR, and HER2), which determine eligibility for targeted therapies such as endocrine therapy, CDK4/6 inhibitors, and anti-HER2 agents. These biomarkers are typically measured using IHC, a test that is typically performed when the tumor’s tissue of origin is known. However, CUP, which accounts for 3-5% of malignancies, lack an identifiable origin and are therefore treated with non-specific chemotherapy, resulting in poor outcomes (1). Notably, a subset of CUP tumors likely arise from breast and other ER-, PR-, or HER2-expressing tissues of origin, and may therefore be amenable to targeted therapies. Yet without a known primary site, receptor profiling is rarely performed. As a result, tumor receptor status remains unrecognized, and effective targeted therapies go unprescribed.

Beyond CUP, standard IHC introduces an additional limitation. Clinical practice often reduces receptor expression to discretized categories, commonly binarized as positive or negative, discarding prognostically meaningful variation along a biological continuum (2; 3; 4). Borderline cases are poorly served by such hard thresholds, and variability in IHC staining and interpretation introduces noise into the label itself. Continuous probabilistic outputs may offer a more faithful representation of this spectrum, enabling uncertainty-aware risk stratification that binary calls cannot provide.

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

This work. We introduce MESH-HR, a multimodal approach that predicts receptor subtype probabilities from

H&E WSIs and structured somatic genomic features available in CUP. MESH-HR integrates an ABMIL image encoder with an XGBoost model over structured genomics, capturing complementary biological signals across modalities. Rather than discretizing receptor status as in standard clinical practice, it outputs calibrated continuous probabilities. We demonstrate improved performance over unimodal baselines, enhanced survival stratification using continuous outputs, and biologically consistent receptor predictions in CUP.

2. Method

Data and labels. We assembled 1,368 breast cancer cases (cancers of known primary; CKPs) from an institutional targeted sequencing cohort, each with a paired H&E whole-slide image and somatic genomic profile. Clinical receptor labels (ER, PR, HER2) were extracted from unstructured pathology reports using GPT-4 and used as ground truth; a random subset was manually reviewed to verify extraction accuracy. We additionally evaluated zero-shot generalization on 786 CUP cases from the same cohort, for which no IHC receptor labels were available; tumor lineage was inferred using an independent tissue-of-origin model (withheld for review). External validation was performed on TCGA-BRCA, a public breast cancer cohort comprising paired histopathology and genomic data.

Vision branch. WSIs were processed with TRIDENT (8; 9) for tissue segmentation and 256×256 patch extraction. Patch embeddings were generated using ResNet50 and aggregated into slide-level predictions via a gated ABMIL model (10; 11).

Genomics branch. The genomics branch uses gene-level somatic mutation indicators, copy number alteration (CNA) scores, and mutational signature weights as input features. An XGBoost classifier (12) was used for prediction.

Fusion. Predictions from both branches are combined via an equal-weight ensemble: the arithmetic mean of ABMIL and XGBoost predicted probabilities. MESH-HR outputs both a continuous probability and a thresholded binary call per receptor.

Training and evaluation. Independent models were trained for each receptor and tuned using randomized hyperparameter search within five-fold nested cross-validation; final performance was evaluated on a held-out test set.

3. Experiments

3.1. Multimodal Integration Outperforms Unimodal State of the Art Across Internal and External Cohorts

MESH-HR consistently achieves the highest AUC across all three receptors on the internal held-out test set (ER 0.90, PR 0.84, HER2 0.96), outperforming both unimodal baselines by up to 0.10 AUC points for ER, 0.08 for PR, and 0.23 for HER2 (Table 1). 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 were enriched for tumor tissue over stroma; in the genomics branch, HER2 prediction is driven by *ERBB2* CNA (encoding HER2) and ER prediction by *TP53* mutation, while *ESR1* (encoding ER) ranks low in feature importance, consistent with its role in metastatic rather than primary disease.

Applied without retraining to TCGA-BRCA (13), 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. These results exceed published 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. (14) achieve HER2 AUC 0.81 (zero-shot). MESH-HR exceeds all three baselines despite training on an independent cohort, predicting all receptors simultaneously.

To benchmark against prior work, we compare MESH-HR to EXPAND (7), an imaging-only method we evaluate on our internal cohort that classifies H&E WSIs into HR+ (ER or PR positive), HER2+, and triple-negative breast cancer (TNBC; ER, PR, and HER2 negative) using nucleus-level features. EXPAND outputs discrete labels and does not produce receptor-level probabilities; to match this setting, MESH-HR predictions are thresholded at 0.5 and mapped to the same phenotype definitions. Predictions are generated on half of the dataset using out-of-fold estimates to increase statistical power without leakage. Because evaluation is performed on discrete multi-class labels rather than continuous

Table 1. AUC for per-receptor classification on the internal held-out test set and external TCGA-BRCA cohort under zero-shot evaluation (no retraining).

Model	Cohort	ER	PR	HER2
ABMIL (H&E only)	Internal	0.80	0.76	0.73
XGBoost (genomics)	Internal	0.82	0.80	0.95
MESH-HR	Internal	0.90	0.84	0.96
MESH-HR	TCGA-BRCA	0.90	0.91	0.94

Table 2. Comparison of MESH-HR and EXPAND on HR+, HER2+, and TNBC classification in the internal cohort. Balanced accuracy and macro F1 are reported due to class imbalance.

Phenotype	Model	Bal. Acc	Macro F1
HER2+	MESH-HR	0.91	0.94
HER2+	EXPAND	0.50	0.48
HR+	MESH-HR	0.74	0.76
HR+	EXPAND	0.50	0.23
TNBC	MESH-HR	0.65	0.69
TNBC	EXPAND	0.50	0.08

scores, we report balanced accuracy and macro F1 (Table 2).

EXPAND achieves balanced accuracy of 0.50 across all subtypes (HER2+, HR+, TNBC), equivalent to random chance, whereas MESH-HR achieves 0.91, 0.74, and 0.65, with macro F1 scores of 0.94, 0.76, and 0.69. This likely reflects the limited expressivity of EXPAND’s interpretable, nuclear morphology–based features under domain shift, which restrict the model to a single cellular compartment. In contrast, MESH-HR leverages whole-slide tissue context and complementary genomic features, enabling substantially improved performance.

3.2. Continuous Predictions Improve Prognostic Signal Beyond Binary and Clinical Labels

Beyond classification, MESH-HR’s multimodal predictions can serve as digital biomarkers of clinical outcome. Breast cancer survival is known to differ by HR status, both because ER/PR+ tumors are biologically less aggressive (15) and because ER/PR+ and HER2+ tumors are amenable to targeted therapies. We compared MESH-HR classifications to clinical HR labels, which may 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 evaluated them for association with overall survival using a multivariable Cox proportional hazards model adjusted for age, tumor purity, biopsy site and type, and sequencing panel version.

Including both binary MESH-HR HR status (posterior probability ≥ 0.5) and binary clinical HR label as covariates, MESH-HR is a stronger predictor of overall survival (hazard ratio = 0.572, $p = 4.54 \times 10^{-6}$) than the clinical label (hazard ratio = 0.717, $p = 0.003$) (Figure 1A), with HR-positive status protective in both cases. The stronger effect size of MESH-HR suggests the model recovers a prognostic signal that more closely reflects true receptor biology than the clinical label alone.

Continuous MESH-HR scores further improve over binary predictions. In a joint Cox model including both continuous and binarized ER and PR predictions, only the con-

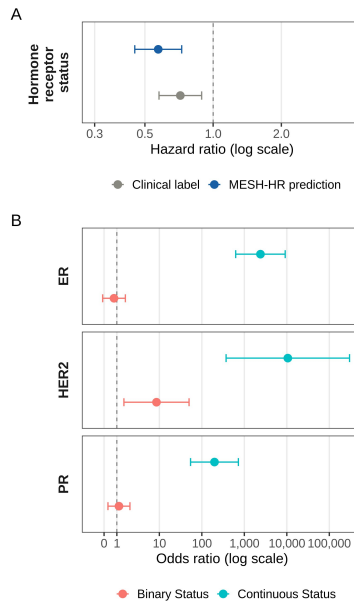


Figure 1. Forest plots showing (A) hazard ratios from multivariable Cox proportional hazards models for MESH-HR predictions and clinical IHC labels as predictors of overall survival in breast CKPs, and (B) odds ratios from multivariable logistic regression models for MESH-HR–derived continuous and binary receptor predictions against clinical ER, PR, and HER2 labels.

tinuous ER score remains significant (hazard ratio = 0.29, $p = 0.0012$); all other terms are non-significant (continuous PR $p = 0.32$, binary ER $p = 0.94$, binary PR $p = 0.79$). Similarly, in logistic regression models predicting clinical receptor labels from both continuous and binarized MESH-HR outputs, continuous scores show substantially stronger associations (ER $p = 5.99 \times 10^{-30}$, PR $p = 1.26 \times 10^{-15}$, HER2 $p = 5.66 \times 10^{-8}$), while binarized predictions show weak or non-significant associations (ER $p = 0.63$, PR $p = 0.71$, HER2 $p = 0.017$) (Figure 1B). These results indicate that binary thresholding discards meaningful prognostic signal, whereas continuous probabilistic predictions preserve it.

3.3. Zero-Shot Receptor Prediction in 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 a known tissue of origin and often lack reliable receptor IHC measurements. Although trained on breast cancer, MESH-HR may remain informative in CUP due to the presence of breast-origin cases and other tissue lineages capable of expressing hormone receptors. To contextualize predictions, we used an independent tissue-of-origin model (withheld for review), to infer tumor lineage.

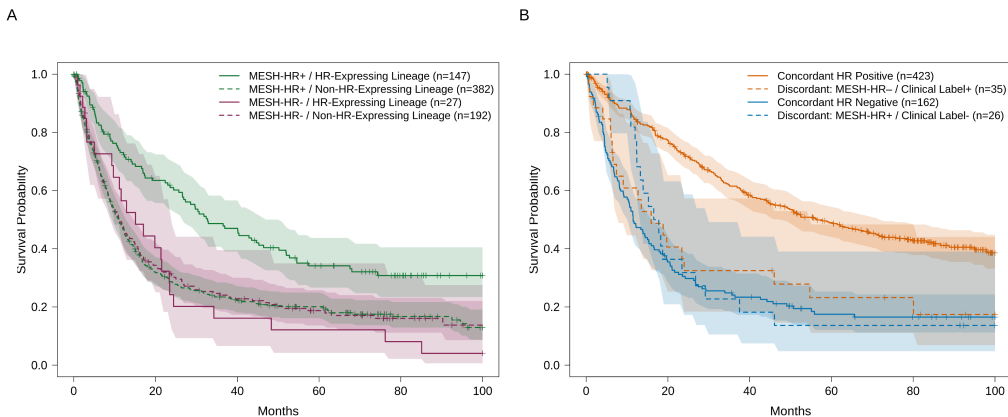


Figure 2. Kaplan–Meier survival curves for (A) CUP patients stratified by MESH-HR predicted HR status and model-inferred tissue lineage (putatively HR-expressing vs. non–HR-expressing), and (B) breast CKPs stratified by concordance between MESH-HR predicted HR status and clinical HR labels. Groups in (A) correspond to HR-positive and HR-negative predictions within each lineage; groups in (B) correspond to concordant and discordant label–prediction pairs. MESH-HR+: predicted HR-positive; MESH-HR–: HR-negative.

Among CUPs with predicted breast origin or lineage, MESH-HR recapitulates the canonical survival advantage of HR-positive disease observed in breast cancer. Predicted HR-positive status is strongly associated with improved overall survival (hazard ratio = 0.04, $p = 4.9 \times 10^{-4}$), despite the absence of receptor labels.

Across all CUPs, MESH-HR also recovers the expected relationship between receptor expression and tissue context: predicted receptor probabilities are significantly higher in lineage-compatible tissues than in incompatible lineages for ER, PR, and HER2 (all $p \ll 0.001$), indicating biologically meaningful generalization beyond breast cancer. A multivariable Cox model with an interaction between predicted HR status and hormone-expressing lineage shows that the survival advantage of HR-positive predictions is present only in receptor-compatible tissues and absent in incompatible lineages (interaction hazard ratio = 0.49, $p = 0.005$), where survival is uniformly poor regardless of predicted HR status (Figure 2A).

Together, these results demonstrate that MESH-HR generalizes zero-shot to CUP and recovers biologically and clinically meaningful survival structure, enabling receptor-informed stratification without access to tissue of origin or IHC measurements.

3.4. Discordant Cases Follow Model-Predicted Survival Profiles

To assess whether MESH-HR captures tumor biology beyond clinical labels, we stratified breast CKPs into four groups based on agreement between MESH-HR predictions and clinical HR status: concordant HR+ (MESH-HR+/label+), discordant MESH-HR–/label+, concordant HR– (MESH-HR–/label–), and discordant MESH-

HR+/label–.

Discordant MESH-HR–/label+ cases have markedly worse survival than concordant HR+ cases, approaching the survival of HR– disease (hazard ratio = 2.02, $p = 0.005$; Figure 2B). Conversely, discordant MESH-HR+/label– cases show improved survival relative to concordant HR– cases (hazard ratio = 0.79, $p = 0.35$). Across both discordant groups, survival aligns more closely with MESH-HR predictions than with clinical labels, indicating that the model captures biologically meaningful signal not reflected in binary IHC annotations.

4. Conclusion

We present MESH-HR, a multimodal framework that integrates structured somatic genomics with whole-slide histopathology to infer continuous breast cancer receptor profiles without IHC. Across internal and external cohorts, multimodal integration outperforms unimodal baselines, while continuous predictions preserve prognostic signal lost under binary clinical annotations. Model-derived receptor estimates are more predictive of survival than clinical labels and generalize zero-shot to TCGA-BRCA and CUP. These results highlight the complementary value of structured genomics and histopathology, and show that continuous predictions improve prognostic stratification over binary labels.

This work positions structured genomic data as a central modality for multimodal learning in healthcare. Integrating structured somatic features with histopathology improves generalization under label scarcity and domain shift, suggesting structured data can anchor stable, transferable representations. Future work will explore improved fusion, uncertainty calibration, and prospective 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] 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.
- [11] Maximilian Ilse, Jakub M. Tomczak, and Max Welling. Attention-based deep multiple instance learning. *arXiv preprint arXiv:1802.04712*, 2018.
- [12] 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.
- [13] The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*, 490:61–70, 2012.
- [14] 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.
- [15] Harold J. Burstein. Systemic therapy for estrogen receptor-positive, HER2-negative breast cancer. *New England Journal of Medicine*, 383:2557–2570, 2020.