medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

# Corr-A-Net: Interpretable Attention-Based Correlated Feature Learning framework for predicting of HER2 Score in Breast Cancer from H&E Images

Kaushik Dutta,[1,2,*] Debojyoti Pal,[1,2] Suya Li,[1,2] Chandresh Shyam,[2] and Kooresh I. Shoghi [1,2,3,4,**]
[1]Imaging Science Program, Washington University in St Louis, St Louis, MO USA
[2] Mallinckrodt Institute of Radiology, Washington University in St Louis, St Louis, MO USA
[3] Department of Biomedical Engineering, Washington University in St Louis, St Louis, MO USA
[4] Lead contact
*Correspondence: kaushik.dutta@wustl.edu
**Correspondence: shoghik@wustl.edu

## SUMMARY

Human epidermal growth factor receptor 2 (HER2) expression is a critical biomarker for assessing breast cancer (BC) severity and guiding targeted anti-HER2 therapies. The standard method for measuring HER2 expression is manual assessment of IHC slides by pathologists, which is both time intensive and prone to inter- and intra-observer variability. To address these challenges, we developed an interpretable deep-learning pipeline with Correlational Attention Neural Network (Corr-A-Net) to predict HER2 score from H&E images. Each prediction was accompanied with a confidence score generated by the surrogate confidence score estimation network trained using incentivized mechanism. The shared correlated representations generated using the attention mechanism of Corr-A-Net achieved the best predictive accuracy of 0.93 and AUC-ROC of 0.98. Additionally, correlated representations demonstrated the highest mean effective confidence (MEC) score of 0.85 indicating robust confidence level estimation for prediction. The Corr-A-Net can have profound implications in facilitating prediction of HER2 status from H&E images.

## KEYWORDS

HER2 Score, Interpretable AI, Deep Learning, Digital Pathology, Breast Cancer, Feature Learning, H&E

## INTRODUCTION

Breast cancer (BC) is regarded as one of the most widespread cancers worldwide, contributing to the second leading cause of cancer-related mortality in women (1, 2). BC exhibits significant heterogeneity, which underscores the need for conducting pathological assessment for tumor histology to determine optimal treatment planning (3, 4). The human epidermal growth receptor 2 (HER2) oncogene, also known as ERBB2 or neu, encodes the type I receptor tyrosine kinase HER-2/neu promotes cell growth and is typically found in cell membranes. Approximately 15-20% of BC patients demonstrate HER2 gene amplification or overexpression, which is associated with tumor aggressiveness, proliferation and resistance to conventional therapies (5-8). The development of targeted anti-HER2 therapies (acting as HER2 inhibitors), such as using trastuzumab (9), lapatinib (10) and petuzumab (11) has shown to decrease patient mortality and increase outcome (12, 13). In order to administer these treatment strategies accurate measurement of HER2 expression in breast cancer is crucial.

In clinical practice, the HER2 score is assessed by performing Immunohistochemical (IHC) staining, followed by pathologists analyzing tissue samples manually to determine the protein expression level. Hematoxylin and eosin (H&E) counterstaining are employed for the cell nuclei and other tissue features. Machine learning coupled with digital slide scanners have enabled automated analysis of Whole Slide Images (WSIs) for predicting disease characteristics by analyzing complex cellular and protein characteristics and widely applied in breast cancer (14-19). Deep learning methods have the capability to

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

capture multi-scale intricate details from WSIs and find applications in cell segmentation (20), tumor classification (21), biomarker estimation (22, 23) and cancer localization (24). In HER2 score prediction, deep-learning methods are employed at both the slide level and the patch level on immunohistochemistry (IHC) stained slides, enabling a comprehensive analysis of HER2 expression and enhancing predictive accuracy (25-29). H&E slides are also used extensively to predict HER2 score leveraging deep-learning pipelines (30-34). Despite these advancements, the absence of biologically interpretable morphological features in H&E-stained slides limits their utility for the quantitative prediction of HER2 expression levels (35). In recent times, the development of Correlational Neural Network (CorrNet) (36) has shown capabilities to produce common representations from multi-modal data by leveraging concept of canonical correlation and multi-modal autoencoder from unimodal data post training.

In this work we developed a two-step deep-learning based framework to predict the HER2 level expression from H&E slides. The first step employs a Correlational Attention Network (Corr-A-Net) that leverages spatially aligned H&E and IHC slides from breast cancer patients to learn the correlated features between these two imaging modalities. The Corr-A-Net exhibits multi-modal image features fusion using an attention mechanism. In the second step, a predictive network is trained using the correlated representations derived exclusively from H&E images to forecast HER2 expression levels. For the predictive network, we implemented Vision Transformer (37), EfficientNet (38), and ResNeXT-50 (39) architectures, alongside a confidence score predictor that assigns a learnable confidence score to each prediction. Additionally, we leveraged the potential of explainable artificial intelligence by applying Gradient Class Activation Maps (Grad-CAM) (40) to identify the morphological features contributing to the predictive decisions. The overview of the framework is given in Figure. 1.

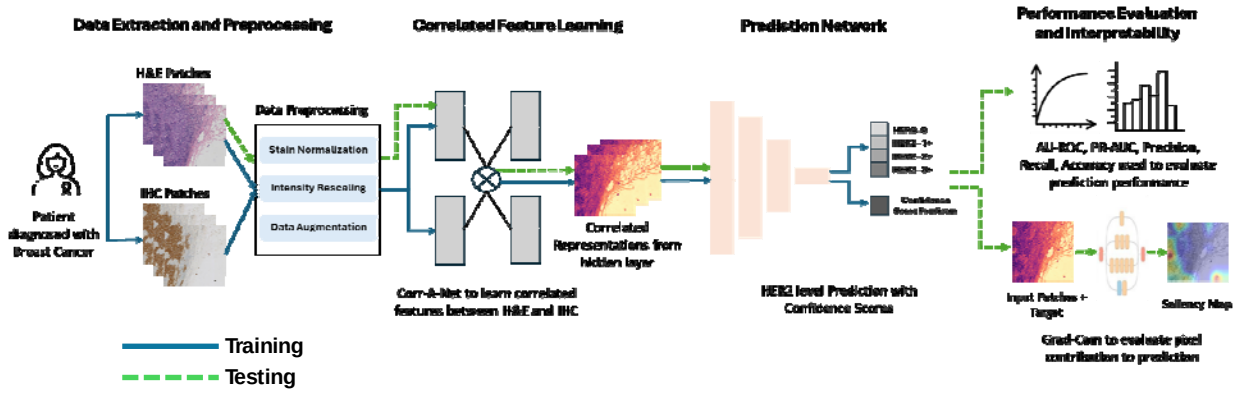



**Figure 1.** Overview of the proposed framework to predict HER2 score using correlated feature learning

## RESULTS

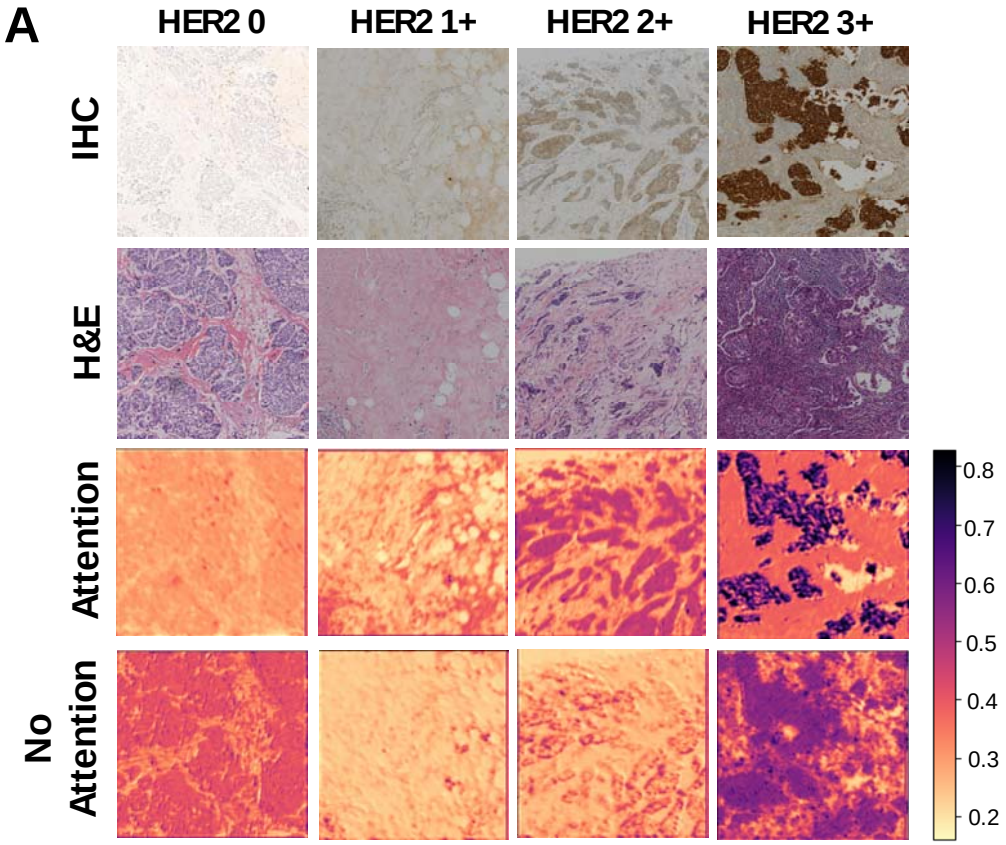
### *Degree of Correlation for Latent Representations*

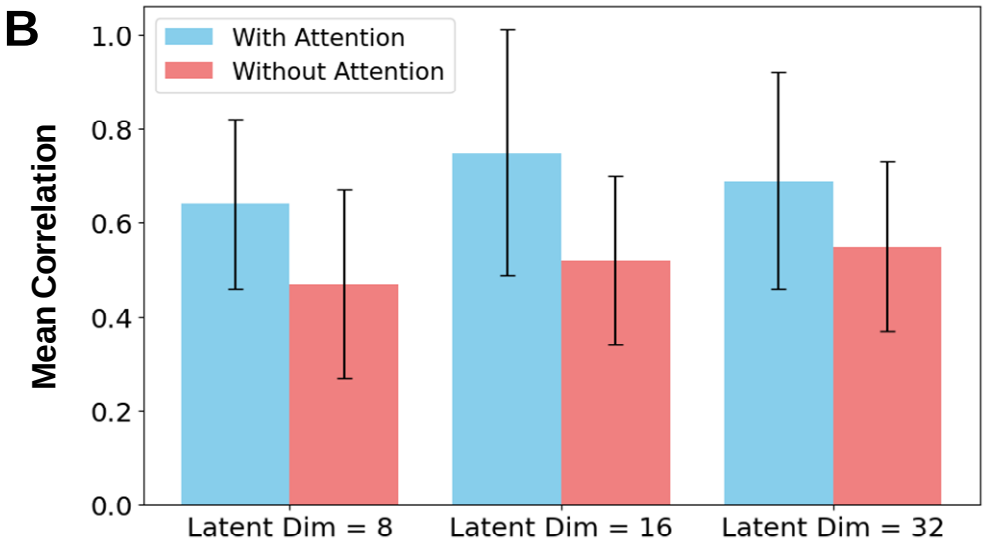
Figure 2A depicts the representative patches from IHC, H&E and hidden/latent representations for different HER2 level expression generated by the Corr-A-Net model. These correlated representations were generated using post-trained Corr-A-Net model from H&E images only. The correlated representations visibly capture important imaging features, particularly the membranous structures characteristic in IHC images which are indicator of HER2 level prediction. This highlights Corr-A-Net's ability to effectively translate imaging features between modalities during training which can be leveraged during inferencing using unimodal data along with giving a dimension of interpretability during prediction.

The ability of Corr-A-Net to learn correlated representations from unimodal data (i.e. H&E data only) is evaluated by computing the mean correlation between the hidden representations and the IHC images. We evaluated the mean correlation performance for two different scenarios of Corr-A-Net i.e. with and

medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

without using the squeeze-excitation attention block and is depicted in Figure 2B. For three different latent dimensions i.e. 8, 16 and 32 we observed that the mean correlation is maximum for dimension = 16 both attention (ρ = 0.76) and dimension = 32 for without attention (ρ = 0.55) scenario. The shared representations using the SA block (i.e. using attention mechanism) in the Corr-A-Net exhibited significantly better (*p≤0.05, Wilcoxon Test*) mean correlation score compared to without using the SA block.



medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

**Figure 2. (A)** Correlated Representations **(B)** Mean Correlation between the hidden representations and the IHC images for different filter sizes (8, 16, 32) of Corr-A-Net for both with and without attention scenarios

### *HER2 level Prediction Performance*

The HER2 level prediction performance using of the deep-learning based classifiers (ViT, EfficientNet, ResNeXT) were evaluated using BA, WP, WR and MCC on the independent testing dataset (978 patches). The performance of the correlated representations obtained from H&E only in predicting HER2 level was benchmarked against classification using H&E slides alone, IHC slides alone, as well as combined IHC and H&E slide inputs, providing a comparative analysis of modality-specific and multimodal prediction approaches. The prediction performance of the classifiers is summarized in Table. 1.

**Table. 1.** HER2 Expression Level Prediction Performance results using WP, WR, BA, MCC, AUC and PR-AUC for different models.

| | Model | WP | WR | BA | MCC | AUC | PR-AUC |
|---|---|---|---|---|---|---|---|
| **H&E Only** | EfficientNet | 0.69 | 0.64 | 0.60 | 0.64 | 0.77 | 0.70 |
| | ViT | 0.72 | 0.75 | 0.72 | 0.65 | 0.82 | 0.75 |
| | ResNeXT | 0.79 | 0.77 | 0.80 | 0.77 | 0.91 | 0.84 |
| **IHC Only** | EfficientNet | 0.74 | 0.73 | 0.63 | 0.66 | 0.83 | 0.76 |
| | ViT | 0.82 | 0.80 | 0.76 | 0.71 | 0.88 | 0.80 |
| | ResNeXT | 0.87 | 0.88 | 0.85 | 0.81 | 0.94 | 0.91 |
| **H&E and IHC Combined** | EfficientNet | 0.87 | 0.86 | 0.87 | 0.80 | 0.84 | 0.73 |
| | ViT | 0.88 | 0.87 | 0.88 | 0.82 | 0.94 | 0.88 |
| | ResNeXT | 0.91 | 0.90 | 0.92 | 0.86 | 0.98 | 0.96 |
| **Corr-A-Net (Filter Size = 8) (H&E Only)** | EfficientNet | 0.86 | 0.85 | 0.84 | 0.78 | 0.84 | 0.72 |
| | ViT | 0.86 | 0.86 | 0.87 | 0.79 | 0.95 | 0.89 |
| | ResNeXT | 0.88 | 0.88 | 0.89 | 0.83 | 0.97 | 0.94 |
| **Corr-A-Net (Filter Size = 16) (H&E Only)** | EfficientNet | 0.88 | 0.87 | 0.83 | 0.82 | 0.88 | 0.76 |
| | ViT | 0.86 | 0.85 | 0.89 | 0.78 | 0.95 | 0.9 |
| | ResNeXT | 0.91 | 0.90 | 0.94 | 0.87 | 0.98 | 0.95 |
| **Corr-A-Net (Filter Size = 32) (H&E Only)** | EfficientNet | 0.87 | 0.87 | 0.84 | 0.81 | 0.88 | 0.75 |
| | ViT | 0.86 | 0.87 | 0.89 | 0.81 | 0.94 | 0.88 |
| | ResNeXT | 0.93 | 0.92 | 0.93 | 0.90 | 0.97 | 0.96 |

medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint
**(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

The HER2 level prediction using H&E alone exhibited the lowest performance across all evaluated network models, achieving a BA score of 0.60, 0.72 and 0.80 for EfficientNet, ViT and ResNeXT respectively. The IHC alone achieved a BA score of 0.63, 0.76, and 0.85 for EfficientNet, ViT and ResNeXT respectively. However, when the IHC patches were integrated with H&E patches, the performance improved significantly, and the BA increased by 0.24, 0.12, and 0.07 for EfficientNet, ViT, and ResNeXT, respectively, relative to the IHC-only approach. These results demonstrate that multimodal integration of H&E and IHC patches significantly improves classification performance compared to unimodal data. The correlated representations generated by Corr-A-Net by utilizing H&E patches alone were assessed for predicting HER2 level expression across three distinct latent dimensions i.e. 8,16 and 32. In all cases, the representations exhibited significant improvement in HER2 level prediction compared to the IHC only model. Furthermore, when compared to the performance of models leveraging combined H&E and IHC patches, the correlated representations yielded a marginally better BA score for ResNeXT and ViT models using latent dimensions of 16 and 32. Overall, the ResNeXT consistently demonstrated superior BA scores compared to other classifier models across all evaluated latent dimensions.

medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

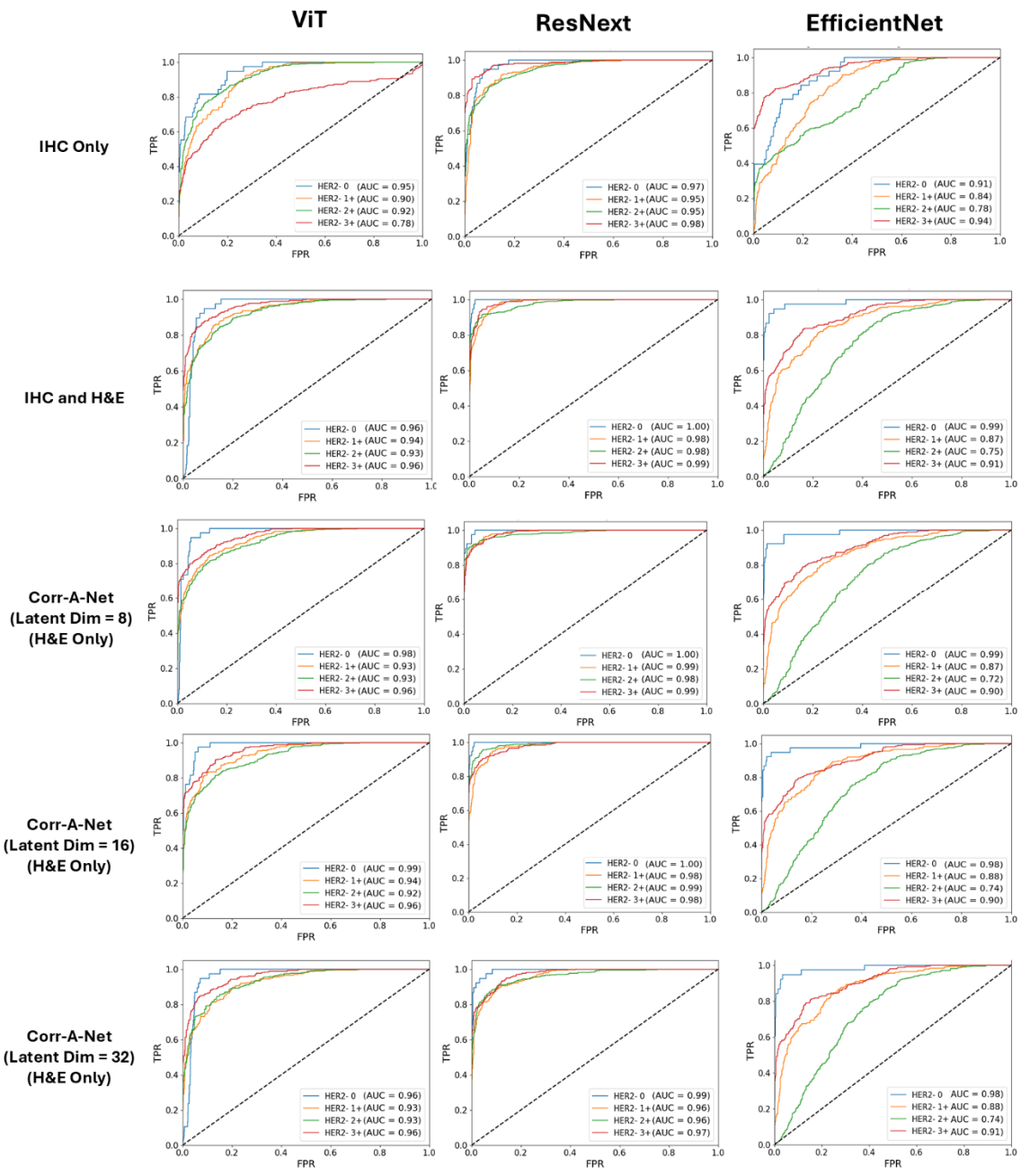



**Figure 3.** ROC curve and AUC values for EfficientNet, ResNeXT and ViT stratified for different HER2- level expression for IHC only, IHC and H&E combined and correlated representations.

Figure. 3 depicts the Receiver Operating Characteristics (ROC) curve and the AUC values for IHC only, IHC and H&E combined and correlated representations in predicting HER2 level. The class-wise ROC was plotted to determine class-wise discriminative capability of the models for different scenarios. From Figure. 5 it is evident that for all network model the IHC and H&E combined, and the correlated representations AUC performance are better for all different classes compared to the IHC only model.

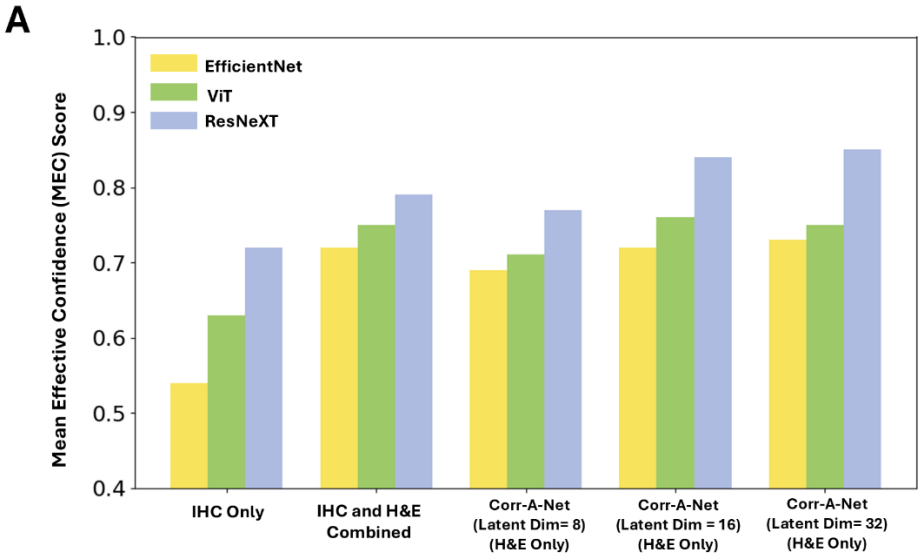
### *Confidence Score Evaluation*

The accuracy of estimating confidence scores associated with HER2 prediction by the models is evaluated using the MEC metric. Figure 4A illustrates that predictions derived from correlated

medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

representations and those based on combined IHC-H&E inputs exhibited significantly improved MEC scores compared to predictions derived solely from IHC data. Particularly, ResNeXT based classification using correlated representations for latent dimensions of 16 and 32 exhibited the best MEC scores of 0.84 and 0.85 respectively. Consistent with the HER2 level prediction performance ResNeXT showed better MEC scores compared to its peer network models for all different settings.

Generally, higher MEC indicates that a greater proportion of correct predictions were made with high confidence score, while incorrect predictions were made with lower confidence score. The relationship of classification performance to the estimated confidence scores were evaluated by performing a stratification of patches based on the confidence score. Three level of stratification level were chosen i.e. for Confidence Score (CS) ≥ 0.9 (high confidence), 0.75 ≤ CS < 0.9 (medium confidence), CS < 0.75 (low confidence) and the classification performance was evaluated for all stratification level using ROC analysis (Figure. 4B). The AUC score was consistently higher for all the scenarios when the CS was greater than 0.9 suggesting that the confidence predictor network consistently predicts correct confidence level for correct predictions.



medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

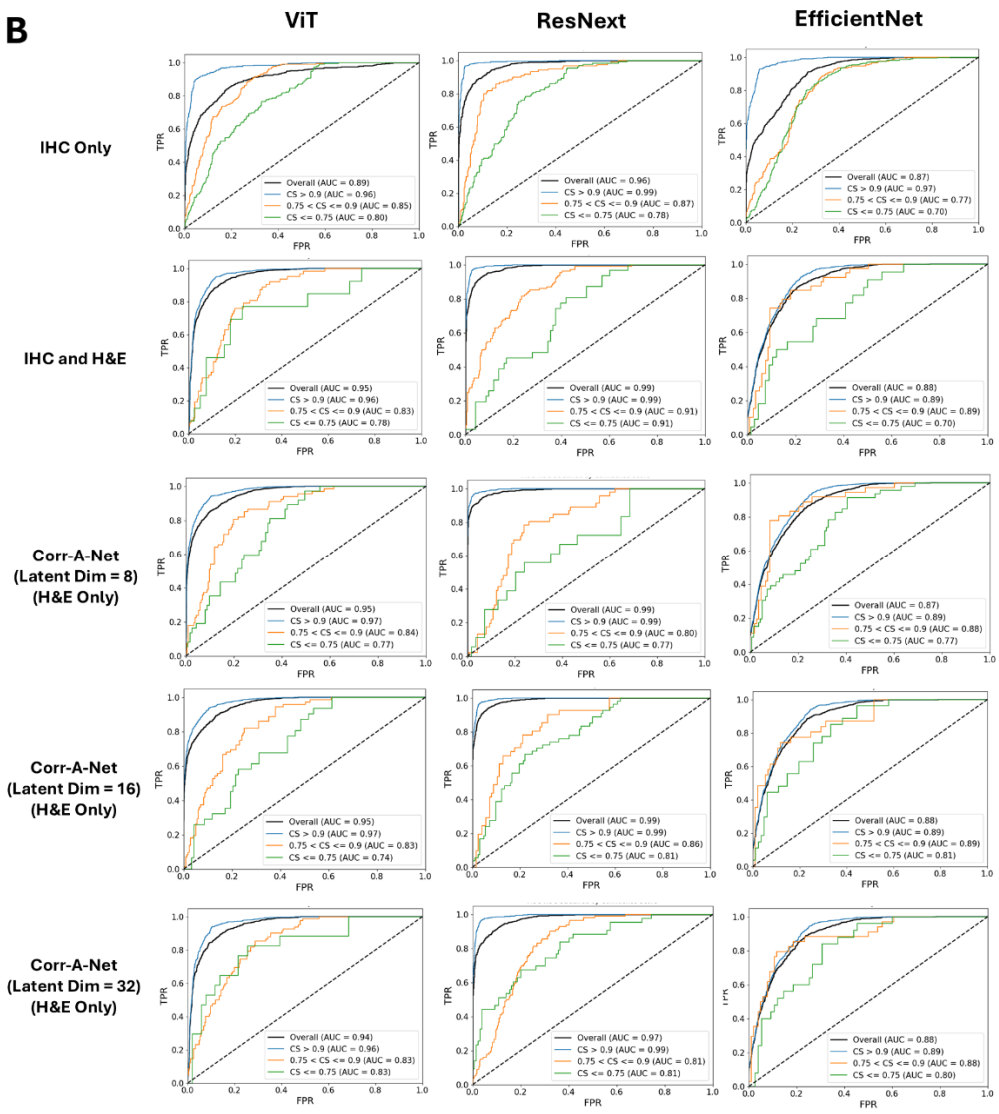



**Figure 4. (A)** Bar graph depicting the MEC score **(B)** ROC curve and AUC values for EfficientNet, ResNeXT and ViT stratified by different Confidence Score (CS)

## *Explainability Analysis*

Figure 5A depicts the Grad-CAM map to assess interpretability of the prediction network in two different scenarios i.e. when using correlated representations subjected to attention block and when not subjected to attention block. The visual inspection of the Grad-CAM maps suggests that attention based correlated representations have better ability of localizing the relevant areas for prediction when compared to the no-attention based correlated representations. The Grad-CAM's performance in generating relevant explanation maps were evaluated using the Drop in Accuracy metric as shown in Figure 5B. The predictions derived from attention based correlated representations had consistently lower percent of Drop in Accuracy for all the latent dimensions compared to the no-attention scenario. The most effective explanation maps were obtained utilizing ResNeXT model with latent dimension of 16, yielding a drop in accuracy of only 20.32%, while the ViT model for the same configuration achieved slightly higher drop of 22.24%. This suggests that attention mechanism can effectively guide the model to focus on regions which have higher likelihood of contributing to accurate predictions.

medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

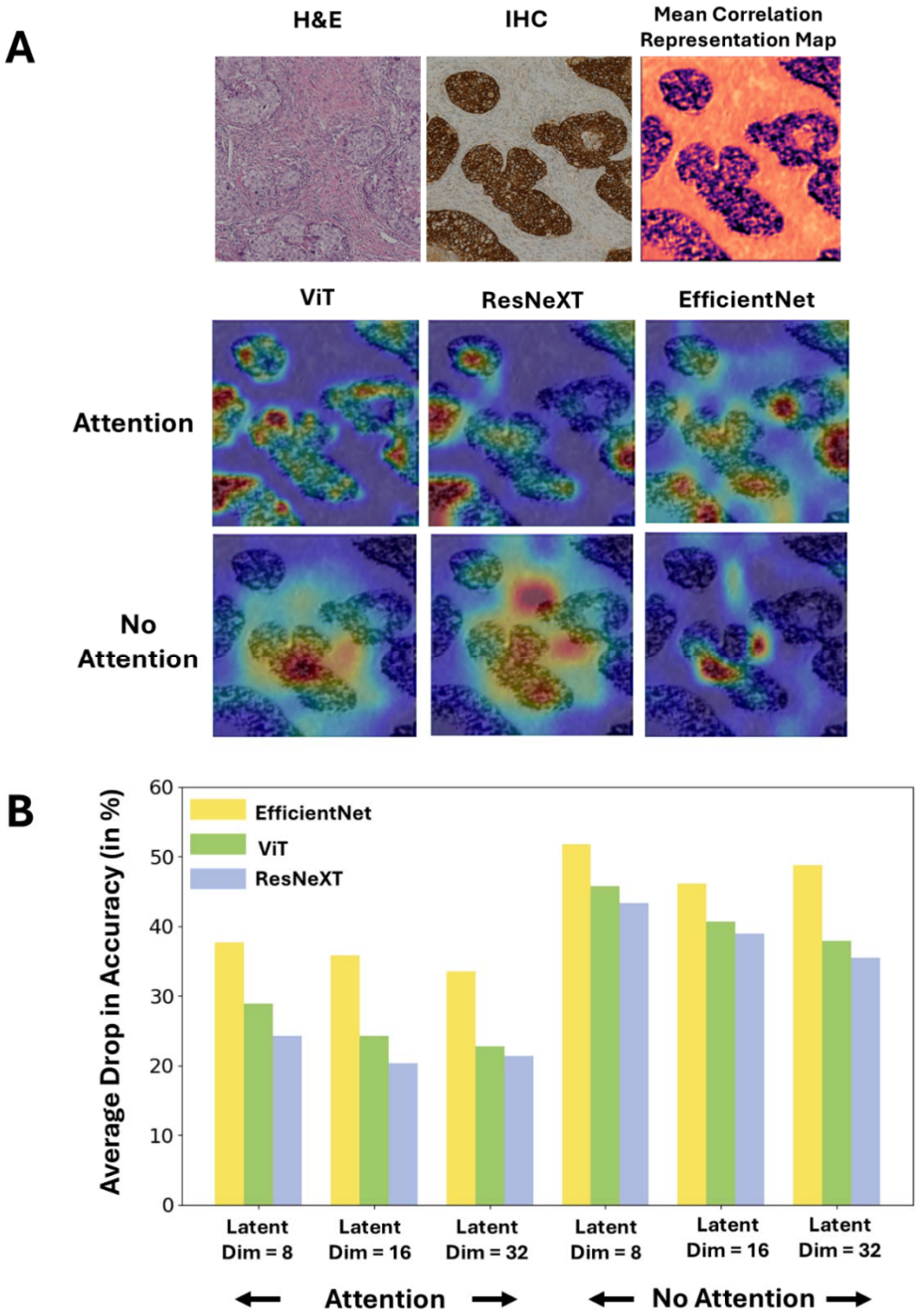



**Figure 5. (A)** Explainability maps obtained using Grad-CAM for ViT, ResNeXT and EfficientNet using attention based and no-attention based correlated representation **(B)** Bar-Graph depicting the Drop in Accuracy (in %) for different prediction network under attention and no-attention scenario.

## DISCUSSION

The objective of this work was to develop a framework to predict the HER2 expression score by utilizing only H&E images. To that end, we developed a novel Corr-A-Net architecture which is trained using IHC and H&E

medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

images to generate correlated representations. Though Corr-A-Net was trained jointly on both modalities, it is capable of generating correlated representations from H&E images alone in post-training phase. Importantly, the correlated representations generated by Corr-A-Net incorporate IHC-based features, thereby providing biological significance to the framework, as demonstrated by our interpretability analysis. The H&E derived correlated representations are subsequently leveraged by three distinct classifier networks to predict HER2 level expression. Additionally, our developed framework includes a confidence prediction module that aids in stratification of cases and enhances interpretability. Previous state-of-the-art approaches for HER2 prediction from H&E images included HAHNet (32), which employed a parallel InceptionV3 architecture with channel attention blocks, and HE-HER2Net (41), a transfer learning-based model utilizing XceptionNet. In comparison, our pipeline achieved the highest overall ROC-AUC score of 0.98 and a balanced accuracy of 94.26% for ResNeXT (latent-dim = 16) in predicting HER2 status from derived correlated representations, which is a 7.5% and 1.5% improvement over HE-HER2Net and HAHNet performance. Moreover, these networks did not incorporate the confidence predictor component or the interpretability analysis to signify biological relevance.

Clinically, HER2 status is assessed using IHC images; however deep-learning based methods have utilized either IHC or H&E separately to predict HER2 scores (15, 16, 21, 25, 26, 28, 30-35). However, H&E images exhibit distinct morphological features compared to IHC and do not capture the HER2 protein expression which is essential for clinical interpretation of HER2 level prediction. For this reason, we trained the Corr-A-Net which effectively learns the correlated feature space between both modalities such that post-training it can be leveraged to generate correlated representations from unimodal data (H&E only). Recent efforts have established that common representation learning through correlational neural networks is capable of generating correlated representation between histopathology (H&E) and radiology (MRI) using MRI data only, facilitating cancer detection/segmentation and interpretability (42, 43). The integration of the attention mechanism in the Corr-A-Net enables the model to compute weighted feature maps by selectively emphasizing critical spatial features while supressing less informative ones. The attention mechanism included within CNN model has demonstrated to produce better classification results for natural (44) and medical images (45, 46) which was persistent in our study as the correlated representations from attention based Corr-A-Net outperformed the no-attention based representations in terms of predictive performance.

Most HER2 prediction algorithm evaluation rely on traditional classification metrics, undermining the quantitative reliability of the prediction. Confidence scores, which quantifies the probability of predictive correctness have been extensively used object detection/segmentation though frameworks like R-CNN and YOLO (47-49). In image classification, the distribution and entropy of the class probabilities are utilized to derive confidence estimates (50, 51). Recent advancements in pathology image analysis have applied confidence estimation through using Multiple Instance Learning (MIL) (52) and ensemble models (53). In MIL, the confidence scores are computed from the probability of an instance belonging to a specific bag of classes, while in ensemble models, confidence is calculated as the mean over precision scores for varying thresholds applied across multiple models. Borquez et al. implemented a Bayesian deep-learning classifier for predicting HER2 expression with Monte-Carlo dropout to simulate multiple classification outputs which are leveraged to model epistemic uncertainty and confidence in prediction (27). While these methods yield a confidence estimate, they are not learnable, meaning that the model predictions are not leveraged as feedback to further enhance performance. Wan et. al developed a confidence network (ConfNet) integrated within the CNN framework which automatically learns the confidence score in the training phase by utilizing negative correlation between model accuracy and task-loss (54). Building on this, we developed an incentivized mechanism for training our prediction models which incorporates a surrogate confidence network. The network learns to estimate confidence dynamically from prediction model's performance, thus enabling learnable confidence. Our results demonstrated in Fig. 4 validates that low confidence is associated with lower AUC and lead to higher rate of misclassification, indicating that model exhibits more confidence for accurate predictions.

The inherent "black-box" nature coupled with limited interpretability and complexity of deep-learning based models limits their translation to clinical practices, as clinicians struggle to trust the reasoning of decision making (55, 56). Explainable artificial intelligence (XAI) which aims to provide task-based interpretation to model outcome has been popularized for medical image classification with the advent of class-activation maps (CAM) based techniques (57, 58). Interpretability is particularly critical in predicting HER2 level expression, as pathologists rely on assessing the intensity of protein expression within the membranous structure of IHC images. Previous work in HER2 level prediction have explored interpretability by analysing the spatial

medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

distribution of attention scores from MIL (59) and by computing the HER2 positive tumour cell percentage (60). Huang et. al have demonstrated that Grad-CAM ++ method revealed stronger influence in diseased tissue areas while predicting HER2 level from ultrasound images (61). These methods did not quantify how predictive performance changes with change in network model which is addressed in our framework by the utilization of drop in accuracy metric. Additionally, attention mechanism (by applying SA-block) has proven to further improve interpretability by guiding the model to focus on highly predictive markers during decision making (44). Our pipeline validated this effect in context of HER2 level prediction where the attention mechanism generated more informative interpretable maps with Grad-CAM than its non-attention counterpart (Figure 5).

Although the correlated representations generated from H&E showed promising and interpretable performance in predicting HER2 status, the framework has few limitations. First, the framework was trained and evaluated on a dataset obtained from the same slide scanner and from the same institution. It is well established that pathological patches vary significantly in terms of slide intensity and quality due to variability in acquisition protocols (62, 63). Although our algorithm incorporates intensity normalization and compute correlated representations in feature space to reduce slide specific variability. a comprehensive validation on external dataset containing pathological specimens acquired under different settings would enhance the generalizability of the network. Second, an ensemble-based approach rather than single network can further improve the predictive performance of the pipeline but at the cost of higher computation and training complexities. Finally, the explainability maps of the Grad-CAM method should be validated by manual pathologist delineation of predictive regions to make it more robust for clinical translation.

## METHODS

### *Dataset and Preprocessing*

We utilized the publicly available Breast Cancer Immunohistochemical (BCI) dataset (64) which consists of total 4873 spatially co-registered H&E and IHC image patches with a size of 1024 × 1024. The dataset was obtained using Hamamatsu NanoZommer S60 pathology section scanner at a scanning resolution of 0.46 µm per pixel from a total of 319 breast cancer patients. Each image patch was labelled for their respective HER2 level expression (0, 1+, 2+ and 3+). The interpretation of each HER2 level expression is given in Table. 2 and representative slide patches from the BCI datasets are depicted in Figure. 6. The dataset was split into training and testing with 3896 samples for training and 977 independent samples for testing the performance. The training samples were further split into training set and validation using 70:30 split for monitoring the performance of the network during training.

For the data preprocessing step, the image patches were subjected to color normalization using Macenko Normalization to address the variability in staining processes across different tissue samples and slides (65). The image patches were resized to 256 × 256 pixels to reduce the computational processing times and to standardize input dimensions for the pretrained neural network architectures. The image patches intensity level was rescaled such that the intensity values are between 0 and 1. Data augmentation is applied in form of random rotation, intensity scaling, contrast adjustment and flipping to prevent overfitting of the model and increase model generalization.

medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

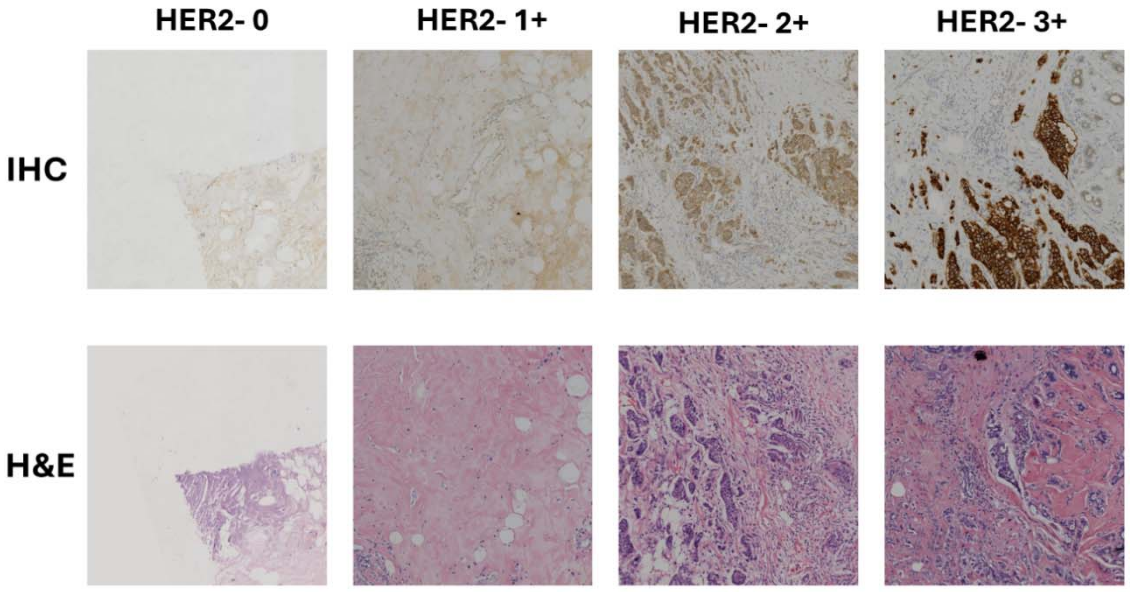



**Figure 6.** Representative slides from the BCI Dataset of H&E and IHC patches for different HER2 expression levels

**Table 2.** HER2 Expression Level and their Interpretation

| HER2 Expression Level/ IHC Score | HER2 Test Interpretation | HER2 Status |
|---|---|---|
| 0 | No Staining or Incomplete and barely perceptible membrane staining in <10% tumor cells | Negative |
| 1+ | Faint and barely perceptible membrane staining in >10% of tumor cells; cells are reactive only in part of their membrane | Negative |
| 2+ | Weak to moderate complete basolateral or lateral membrane staining in ≥ 10% of tumor cells | Equivocal |
| 3+ | Strong and intense complete basolateral or lateral membrane staining in ≥ 10% of tumor cells | Positive |

### *Correlational Neural Network*

We developed a Deep Convolutional Correlational Attention Network (Corr-A-Net) model with three hidden layers having different filter sizes to learn the feature correlation between H&E and IHC patches. We further added a Squeeze-Excitation based attention mechanism in the innermost hidden layer to enhance the correlative feature learning.

*Corr-A-Net Architecture:* The deep convolutional Corr-A-Net is composed of three primary modules: feature extractor (encoder block), feature fusion (hidden layer with attention mechanism) and feature reconstructor (decoder block). The CorrNet is based on the principle of multi-modal autoencoders and canonical correlative analysis and aims to learn a common representation between two input vectors (here obtained from H&E and IHC) such that they are maximally correlated in latent space while at the same time minimizing the self and cross reconstruction loss. In the feature extraction step, the first convolutional block of the ResNet50 architecture (66) was utilized to extract feature maps of size 256 × 256 × 64 from both H&E and IHC patches. This enabled the analysis of correlation between the two modalities in feature space through the use of learned

medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

kernels in the pretrained model. The pixel-level vector representations obtained from H&E patches depicted as $H_i$ and IHC patches depicted as $I_i$ were utilized to train the Corr-A-Net model to learn the common representation. The individual hidden representations $h_H = f(WH_i + b)$ and $h_I = f(VI_i + b)$ are obtained from the multi-layer encoder for each modality for H&E and IHC respectively, where the weights of the modalities are given by $W$ and $V$ for H&E and IHC respectively. The multi-layer approach compared to the single-layer encoder in the original paper (36) enables to learn detailed high-level intricate features from the representations. In our work, we applied three layers of each consisting of series of convolution, batch normalization and non-linear activation using ReLU; where the filter size increased at each layer by a factor of 2. The detailed architecture of the Corr-A-Net used for correlated representation generation is depicted in Figure. 7.

The individual hidden representations i.e. $h_H$ and $h_I$ are then subjected to a squeeze-excitation (SE) attention block to facilitate feature attention mechanism. The SE block aggregates the global spatial information from the modalities by applying a global average pooling operation such that each channel has an attention representation. After the squeeze step the inter-channel dependencies are learned by employing the gating mechanism through the use of fully connected layers followed by non-linear activations (ReLU) and sigmoid function. Finally, the channel-based weightage is multiplied to the original hidden representations of the modalities to get weighted attention maps. The feature fusion module takes the hidden representation from individual modalities and combine to create a shared representation which is subjected to layer of convolution and activation (depicted by $f$) in the innermost hidden layer. The shared representation is given by Eq. 1.

$$h(Z_i) = f(WH_i + VI_i + b) \tag{1}$$

where $Z_i = [H_i, I_i]$ denotes the concatenated vector obtained from the H&E and IHC patches. The shared representation ($h(Z_i)$) is subjected to an identical decoder to obtain the reconstructed view.

*Loss Function and Training of Corr-A-Net:* The Corr-A-Net learns the common representations between the H&E and the IHC image patches by a joint objective function realized by maximizing the correlation between the latent representation of the two modalities (i.e. the correlation loss $L_{Corr}$ ) and by minimizing the reconstruction error of the features from both modalities ($L_{Recon}$). The reconstruction loss is further divided into two parts: self-reconstruction loss and the cross-reconstruction loss. The self-reconstruction loss is calculated by minimizing the error in reconstructing the same modality while the cross reconstruction is calculated by minimizing the error in reconstructing from cross modality. A mixture loss function composed of Structural Similarity Index Metric (SSIM) and L1-loss (Eq. 3) is used as the reconstruction objective where a hyperparameter $\beta$ is used to control the extent between the SSIM and L1-loss (67, 68). The SSIM terms enforces perceptual and structural similarity of the reconstruction view while L1 serves as a regularizer term and helps to maintain the edges. The correlation loss is composed by the latent representations obtained from individual modalities i.e. $h_H$ and $h_I$ and given by Eq. 4. $\overline{h_H}$ and $\overline{h_I}$ represents the mean of the latent representations for H&E and IHC respectively. The hyperparameter $\lambda$ is used to control the extent of correlation loss compared to the reconstruction loss and was optimized to $\lambda = 1.5$ during the training phase.

$$L_{Total} = L_{Recon} - \lambda L_{Corr} \tag{2}$$

$$L_{Recon} = \frac{1}{N}\sum_{i=1}^{N} \beta\left(1 - SSIM\left(Z_i, Z_i^{'}\right)\right) + \left(1-\beta\right)\left(Z_i - Z_i^{'}\right) \tag{3}$$

$$L_{Corr} = \frac{\sum_{i=1}^{N}\left(h_H^i - \overline{h_H}\right)\left(h_I^i - \overline{h_I}\right)}{\sqrt{\sum_{i=1}^{N}\left(h_H^i - \overline{h_H}\right)^2\left(h_I^i - \overline{h_I}\right)^2}} \tag{4}$$

The Corr-A-Net was trained using a gradient based Adam optimizer with adaptive learning rate implemented by the Cosine Annealing algorithm. The network was trained for 250 epochs using two 48GB NVIDIA P8000

medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

GPUs using Pytorch Lightning. After the Corr-A-Net is trained they can be used to generate latent representations without the IHC patches.

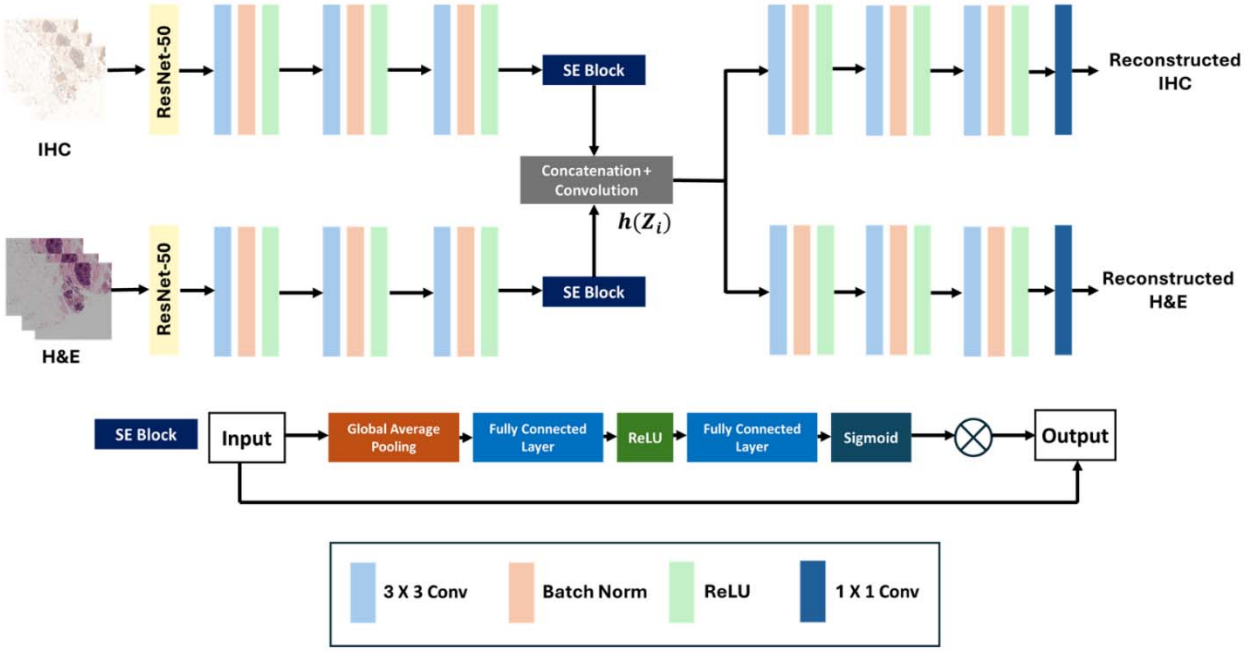



**Figure 7.** Architecture of Correlational Attention Neural Network (Corr-A-Net).

### *Prediction Network and Learnable Confidence Score*

We employed three deep-learning based classification networks—EfficientNet, ResNeXT-50, and Vision Transformer (ViT)—to predict HER2 status using correlated representations derived solely from H&E image patches. In addition to the prediction network, we integrated a confidence estimation network in parallel designed to learn confidence scores of the predictions. The confidence network was trained in an incentivized mechanism encouraging the network to produce reliable predictions. Furthermore, we benchmarked the performance of the correlated representation in predicting HER2 level expression to that of using IHC slides only and using both IHC and H&E combined across the channels.

*Vision Transformer (ViT):* The ViT architecture is based on the transformer architecture which is utilized for image classification task. Unlike, traditional CNNs, ViT segments input images into a sequence of non-overlapping fixed-size patches (16 × 16 pixels) and subsequently projects these patches into learnable linear embeddings with a dimensionality of 1024. To preserve spatial information, learned positional embeddings are incorporated into the linear projections prior to their input into the transformer encoder blocks. The transformer encoder is composed of a series of twelve transformer layers, each employing multi-headed self-attention mechanisms and feed-forward neural networks (multi-layer perceptron) to process the input tokens. Layer normalization and residual connections are integrated within each layer to ensure effective gradient flow during the training process. Furthermore, dropout was applied in the transformer block to differentially switch off neurons to prevent overfitting. Finally, a fully connected layer followed by a SoftMax activation function is applied to produce probability distributions for multi-class classification tasks. This architectural design facilitates the effective modeling of long-range dependencies and complex feature representations inherent in image data.

*EfficientNet:* The EfficientNet architecture is based on the concept of compound scaling method which uniformly scales the network depth, width and resolution with set of scaling parameters. The EfficientNet utilizes a series of Mobile Inverted Bottleneck Convolution (MBConv) blocks with Squeeze-Excitation (SE) modules. In our work we utilized a variant of the EfficientNet B7 model consisting of seven blocks of MBConv

medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

having varying degree of convolution filter size i.e. 3 × 3 and 5 × 5 filters. The MBConv block comprises an expansion phase in which the number of feature filters are increased by the expansion factor, followed by depthwise convolution where convolution is applied separately over each feature channels. The SE block then reduces the feature channels to a single value by applying global-average pulling followed by two fully connected layer with sigmoid activation to generate channel-wise weights which are multiplied to the original feature maps. Finally, the projection layer consisting of 1 × 1 pointwise convolutional blocks converts the feature maps to the same resolution as the input. The use of depthwise convolutional layers increases the time complexity of the network without affecting performance. Finally in the classification head we applied a fully connected dense layer with SoftMax activation function to obtain the multiclass probability.

*ResNeXT:* The ResNeXT architecture is based on the principle of traditional Residual Network (ResNet) model integrated with the concept of cardinality referring to the number of parallel paths within a network block. Each parallel residual sub-block consists of series of convolutions with 3 × 3 filters followed by activation, max-pooling, and feature aggregation by 1 × 1 convolution. In our model we used a cardinality factor of 32 i.e. there are 32 residual sub-blocks which are finally aggregated in the end with the input by using a long degree residual interconnection. Finally, the output is subjected to global average pooling to reduce the dimension followed by a fully connected layer with SoftMax to obtain the multiclass classification probability.

*Confidence Estimation Network:* The fundamental challenge in confidence score estimation lies in the absence of ground truth values for confidence levels. To address this, we applied an incentivized approach in which the model is rewarded to produce confidence estimates which correctly reflects the model's ability to generate correct predictions. The confidence estimation network is coupled with the HER2 score prediction network where the penultimate layer output of the prediction network is subjected to feedforward architecture in parallel to estimate the confidence scores. The feedforward architecture consists of two fully connected layers followed by a sigmoid layer to output a confidence score between 0 and 1 (69). During training in order to give 'hints' to the network the SoftMax prediction probabilities obtained from the HER2 prediction network are calibrated by interpolating between original prediction and ground truth distribution given in Eq. 5.

$$p_i' = c \cdot p_i + (1 - c) \cdot y_i \tag{5}$$

The confidence level ($c$) controls the extent of interpolation when generating the updated predictions given by $p'$. When the model is confident it outputs value such that $c \to 1$ , whereas if the model is unsure of the prediction, it outputs $c \to 0$. We calculated our classification loss by using a weighted cross-entropy function using the updated prediction probabilities ($p'$) and ground truth targets labels ($y$). The weighted cross-entropy loss function addresses class imbalance by assigning higher weights to less frequent classes and lower weights to more prevalent classes and is depicted in Eq. 6.

$$L_{Classification} = -\frac{1}{N}\sum_{i=1}^{N} w_c \, y_{i,c} \log\left(p'_{i,c}\right) \tag{6}$$

where, $w_c$ is the weight vector assigned for different classes. Furthermore, a confidence loss term is applied which acts as a regularizer for the classification loss term and controls the extent of confidence level and is depicted in Eq. 7. Finally, the classification loss is combined with the confidence loss to calculate the final loss function of the network (Eq. 8). The extent of confidence loss is controlled by the hyperparameter $\lambda$.

$$L_{Confidence} = -\frac{1}{N}\sum_{i}^{N} \log c_i \tag{7}$$

$$L_{Total} = L_{Classification} + \lambda \, L_{Confidence} \tag{8}$$

The learnable confidence score affects the dynamics of the loss function i.e. when the confidence score is high the confidence loss term becomes low, and the loss function is subjected to only weighted cross-entropy term. Whereas, when the confidence score is low the confidence loss term is high, and it penalizes the cross-entropy term.

medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

### *Performance Evaluation*

The performance of the models i.e. ViT, EfficientNet and ResNeXT in predicting the HER2 level was evaluated using Weighted Precision (WP), Weighted Recall (WR), Balanced Accuracy (BA) and Matthews Correlation Coefficient (MCC) to account for the imbalanced dataset. To further assess model effectiveness, the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) was calculated, and ROC curves were plotted under varying settings to illustrate the models' discrimination capability. We also computed the Area Under the Precision-Recall Curve (PR-AUC), as this metric is robust in evaluating predictive performance on imbalanced datasets by focusing on precision-recall trade-offs and minimizing potential influence of negative class abundance.

For evaluating the model's ability to estimate the confidence score associated with the prediction we applied the Mean Effective Confidence (MEC) metric depicted by Eq. 9.

$$MEC = \frac{1}{N}\sum_{i=1}^{N} Z_i \cdot c_i \tag{9}$$

$$Z_i = \begin{cases} 1, & when\ i^{th} prediction\ is\ correct \\ -1, & when\ i^{th} prediction\ is\ wrong \end{cases}$$

where $Z_i$ denotes decision parameter whether the prediction is correct or not and $c_i$ is the confidence score associated with that prediction. Thus, if the prediction is correct with high confidence or incorrect prediction with low confidence it leads to higher MEC while if the confidence value is high, but the prediction is incorrect it leads to penalization and decreases the MEC.

### *Explainability Analysis*

The incorporation of explainability in developing CNN models are essential as it helps in visualization of the features which are contributing to the prediction. To this end, we applied the Gradient Class Activation (Grad-CAM) method to generate representative heatmaps to interpret the model decision. We applied the Grad-CAM on the penultimate layer of the prediction network and calculated the gradients flowing backwards from the classification head to the penultimate layer. The gradients indicated the extent to which each activation in the convolutional layer influences the final classification decision. Grad-CAM essentially assigns importance weights to each feature maps based on the calculated gradients by computing the global average across the spatial dimensions. The weights assigned to each feature maps to depict the interpretability is given by Eq. 10.

$$\alpha_k = \frac{1}{N}\sum_i\sum_j \frac{\partial y^c}{\partial A_{i,j}^k} \tag{10}$$

$$L^c = ReLU\left(\sum_k \alpha_k \cdot A^k\right)$$

where, N is the number of pixels, $\alpha_k$ is the weight for the $k^{th}$ feature map, $y^c$ is the target class value, $A_{i,j}^k$ is the gradient activation at position $(i,j)$ for $k^{th}$ feature map. Finally, the activation maps computed for each feature is multiplied with the weights to create the Grad-CAM maps ($L_c$) which signifies the relevance of each spatial location with respect to target class. The effectiveness of Grad-CAM maps lies in its ability to highlight relevant regions important for decision making. To quantitatively assess the performance of Grad-CAM maps in predicting HER2 level, we applied the Drop in Accuracy metric, which is computed as the percentage difference between the model's prediction accuracy on the original image and the model's accuracy when using the Grad-CAM map highlighted regions (Eq. 11). This metric indicates the relevancy of the saliency maps in prediction, with lower drops in accuracy suggesting that Grad-CAM based regions are key features driving the prediction.

$$Drop\ (\%) = \left(\frac{(Accuracy)_{Full-Image} - (Accuracy)_{Grad-CAM}}{(Accuracy)_{Full-Image}}\right) 100\% \tag{11}$$

medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

## RESOURCE AVAILABILITY

### *Lead contact*

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Kooresh I. Shoghi (shoghik@wustl.edu).

### *Materials availability*

This study did not generate new unique reagents.

### *Data and code availability*

- The public dataset used in this study can be downloaded from https://bupt-ai-cz.github.io/BCI/
- All original code has been deposited at Github (Link) and is publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## ACKNOWLEDGMENTS

The authors acknowledge the Center of High-Performance Computing (CHPC) at Mallinckrodt Institute of Radiology, Washington University School of Medicine for computing resources.

## FUNDING

This work was supported by NIH/NCI grants U24CA209837, U24CA253531 and internal funds provided by the Siteman Cancer Center (SCC) and Mallinckrodt Institute of Radiology (MIR). The Center of High-Performance Computing (CHPC) at Washington University in St. Louis for computing resources supported by NIH grant number S10OD025200.

## AUTHOR CONTRIBUTIONS

Conceptualization, K.D.; methodology, K.D., D.P., and S.L.; investigation, K.D., D.P., S.L., C.S., and K.I.S.; writing—original draft, K.D.; writing—review & editing, D.P., S.L., C.S., and K.I.S.; funding acquisition, K.I.S.; resources, K.I.S. and C.S.; supervision, K.I.S.

## DECLARATION OF INTERESTS

The authors have no declaration.

## REFERENCES

1. Giaquinto AN, Sung H, Miller KD, Kramer JL, Newman LA, Minihan A, et al. Breast cancer statistics, 2022. CA: a cancer journal for clinicians. 2022;72(6):524-41.
2. Smolarz B, Nowak AZ, Romanowicz H. Breast Cancer-Epidemiology, Classification, Pathogenesis and Treatment (Review of Literature). Cancers (Basel). 2022;14(10).
3. Zubair M, Wang S, Ali N. Advanced Approaches to Breast Cancer Classification and Diagnosis. Front Pharmacol. 2020;11:632079.
4. Rakha EA, El-Sayed ME, Lee AH, Elston CW, Grainge MJ, Hodi Z, et al. Prognostic significance of Nottingham histologic grade in invasive breast carcinoma. J Clin Oncol. 2008;26(19):3153-8.
5. Ross JS, Slodkowska EA, Symmans WF, Pusztai L, Ravdin PM, Hortobagyi GN. The HER-2 receptor and breast cancer: ten years of targeted anti-HER-2 therapy and personalized medicine. Oncologist. 2009;14(4):320-68.
6. Zhang H, Peng Y. Current Biological, Pathological and Clinical Landscape of HER2-Low Breast Cancer. Cancers (Basel). 2022;15(1).
7. Tarantino P, Hamilton E, Tolaney SM, Cortes J, Morganti S, Ferraro E, et al. HER2-Low Breast Cancer: Pathological and Clinical Landscape. J Clin Oncol. 2020;38(17):1951-62.
8. Perez EA, Cortes J, Gonzalez-Angulo AM, Bartlett JM. HER2 testing: current status and future directions. Cancer Treat Rev. 2014;40(2):276-84.
9. Emens LA. Trastuzumab: targeted therapy for the management of HER-2/neu-overexpressing metastatic breast cancer. American journal of therapeutics. 2005;12(3):243-53.

medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint
**(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

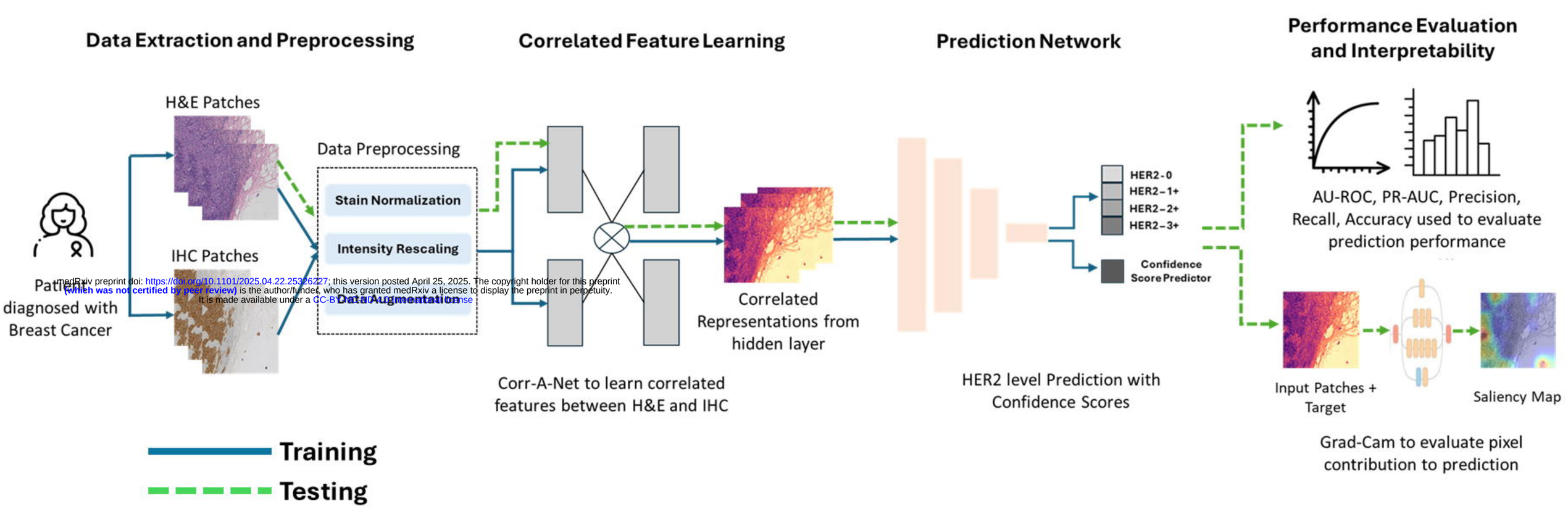
10. Geyer CE, Forster J, Lindquist D, Chan S, Romieu CG, Pienkowski T, et al. Lapatinib plus capecitabine for HER2-positive advanced breast cancer. New England journal of medicine. 2006;355(26):2733-43.
11. Ishii K, Morii N, Yamashiro H. Pertuzumab in the treatment of HER2-positive breast cancer: an evidence-based review of its safety, efficacy, and place in therapy. Core evidence. 2019:51-70.
12. Stanowicka-Grada M, Senkus E. Anti-HER2 drugs for the treatment of advanced HER2 positive breast cancer. Current Treatment Options in Oncology. 2023;24(11):1633-50.
13. Oh DY, Bang YJ. HER2-targeted therapies - a role beyond breast cancer. Nat Rev Clin Oncol. 2020;17(1):33-48.
14. Couture HD, Williams LA, Geradts J, Nyante SJ, Butler EN, Marron J, et al. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. NPJ breast cancer. 2018;4(1):30.
15. Lazard T, Bataillon G, Naylor P, Popova T, Bidard F-C, Stoppa-Lyonnet D, et al. Deep learning identifies morphological patterns of homologous recombination deficiency in luminal breast cancers from whole slide images. Cell Reports Medicine. 2022;3(12).
16. Xue T, Chang H, Ren M, Wang H, Yang Y, Wang B, et al. Deep learning to automatically evaluate HER2 gene amplification status from fluorescence in situ hybridization images. Scientific Reports. 2023;13(1):9746.
17. Cui M, Zhang DY. Artificial intelligence and computational pathology. Laboratory Investigation. 2021;101(4):412-22.
18. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature medicine. 2019;25(8):1301-9.
19. Gamble P, Jaroensri R, Wang H, Tan F, Moran M, Brown T, et al. Determining breast cancer biomarker status and associated morphological features using deep learning. Communications medicine. 2021;1(1):14.
20. Greenwald NF, Miller G, Moen E, Kong A, Kagel A, Dougherty T, et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. Nature biotechnology. 2022;40(4):555-65.
21. Kosaraju S, Park J, Lee H, Yang JW, Kang M. Deep learning-based framework for slide-based histopathological image analysis. Scientific Reports. 2022;12(1):19075.
22. Arslan S, Mehrotra D, Schmidt J, Geraldes A, Singhal S, Hense J, et al. Deep learning can predict multi-omic biomarkers from routine pathology images: A systematic large-scale study. 2022.
23. Niehues JM, Quirke P, West NP, Grabsch HI, van Treeck M, Schirris Y, et al. Generalizable biomarker prediction from cancer pathology slides with self-supervised deep learning: A retrospective multi-centric study. Cell reports Medicine. 2023;4(4).
24. Jiang Y, Huang S, Zhu X, Cheng L, Liu W, Chen Q, et al. Artificial Intelligence Meets Whole Slide Images: Deep Learning Model Shapes an Immune-Hot Tumor and Guides Precision Therapy in Bladder Cancer. Journal of Oncology. 2022;2022(1):8213321.
25. Qaiser T, Mukherjee A, Reddy Pb C, Munugoti SD, Tallam V, Pitkäaho T, et al. Her 2 challenge contest: a detailed assessment of automated her 2 scoring algorithms in whole slide images of breast cancer tissues. Histopathology. 2018;72(2):227-38.
26. Pedraza A, Gonzalez L, Deniz O, Bueno G. Deep Neural Networks for HER2 Grading of Whole Slide Images with Subclasses Levels. Algorithms. 2024;17(3):97.
27. Bórquez S, Pezoa R, Salinas L, Torres CE. Uncertainty estimation in the classification of histopathological images with HER2 overexpression using Monte Carlo Dropout. Biomedical Signal Processing and Control. 2023;85:104864.
28. Saha M, Chakraborty C. Her2Net: A deep framework for semantic segmentation and classification of cell membranes and nuclei in breast cancer evaluation. IEEE Transactions on Image Processing. 2018;27(5):2189-200.
29. Sharma S, Mehra R. Conventional machine learning and deep learning approach for multi-classification of breast cancer histopathology images—a comparative insight. Journal of digital imaging. 2020;33(3):632-54.
30. Farahmand S, Fernandez AI, Ahmed FS, Rimm DL, Chuang JH, Reisenbichler E, et al. Deep learning trained on H&E tumor ROIs predicts HER2 status and Trastuzumab treatment response in HER2+ breast cancer. bioRxiv. 2021:2021.06. 14.448356.
31. Farahmand S, Fernandez AI, Ahmed FS, Rimm DL, Chuang JH, Reisenbichler E, et al. Deep learning trained on hematoxylin and eosin tumor region of Interest predicts HER2 status and trastuzumab treatment response in HER2+ breast cancer. Modern Pathology. 2022;35(1):44-51.
32. Wang J, Zhu X, Chen K, Hao L, Liu Y. HAHNet: a convolutional neural network for HER2 status classification of breast cancer. BMC bioinformatics. 2023;24(1):353.
33. Sekhar A, Goel V, Jain G, Patil A, Gupta RK, Sethi A. HER2 and FISH Status Prediction in Breast Biopsy H&E-Stained Images Using Deep Learning. arXiv preprint arXiv:240813818. 2024.
34. Anand D, Kurian NC, Dhage S, Kumar N, Rane S, Gann PH, et al. Deep learning to estimate human epidermal growth factor receptor 2 status from hematoxylin and eosin-stained breast tissue images. Journal of pathology informatics. 2020;11(1):19.

medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint
**(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

35. Jiao P, Zheng Q, Yang R, Ni X, Wu J, Chen Z, et al. Prediction of HER2 Status Based on Deep Learning in H&E-Stained Histopathology Images of Bladder Cancer. Biomedicines. 2024;12(7):1583.
36. Chandar S, Khapra MM, Larochelle H, Ravindran B. Correlational neural networks. Neural computation. 2016;28(2):257-85.
37. Dosovitskiy A. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:201011929. 2020.
38. Tan M, Le Q, editors. Efficientnet: Rethinking model scaling for convolutional neural networks. International conference on machine learning; 2019: PMLR.
39. Xie S, Girshick R, Dollár P, Tu Z, He K, editors. Aggregated residual transformations for deep neural networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2017.
40. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D, editors. Grad-cam: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE international conference on computer vision; 2017.
41. Shovon MSH, Islam MJ, Nabil MNAK, Molla MM, Jony AI, Mridha M. Strategies for enhancing the multi-stage classification performances of her2 breast cancer from hematoxylin and eosin images. Diagnostics. 2022;12(11):2825.
42. Bhattacharya I, Seetharaman A, Kunder C, Shao W, Chen LC, Soerensen SJ, et al. Selective identification and localization of indolent and aggressive prostate cancers via CorrSigNIA: an MRI-pathology correlation and deep learning framework. Medical image analysis. 2022;75:102288.
43. Bhattacharya I, Seetharaman A, Shao W, Sood R, Kunder CA, Fan RE, et al., editors. Corrsignet: Learning correlated prostate cancer signatures from radiology and pathology images for improved computer aided diagnosis. Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23; 2020: Springer.
44. Bai N, Joe I. Deep Learning Methods with the Improved Attention for Explainable Image Recognition. IEEE Access. 2024.
45. Dutta K, Laforest R, Luo J, Jha AK, Shoghi KI. Deep learning generation of preclinical positron emission tomography (PET) images from low-count PET with task-based performance assessment. Med Phys. 2024;51(6):4324-39.
46. Dutta K, Laforest R, Shoghi K, editors. Self-Supervised learning framework to generate preclinical Standard-Count PET from Low-Count PET and multi-objective task-based performance evaluation. 2023 IEEE Nuclear Science Symposium, Medical Imaging Conference and International Symposium on Room-Temperature Semiconductor Detectors (NSS MIC RTSD); 2023: IEEE.
47. Redmon J, editor You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition; 2016.
48. Wang Y, Wang X, Gao T, Du L, Liu W. An automatic knee osteoarthritis diagnosis method based on deep learning: data from the osteoarthritis initiative. Journal of Healthcare Engineering. 2021;2021(1):5586529.
49. Erhan D, Szegedy C, Toshev A, Anguelov D, editors. Scalable object detection using deep neural networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2014.
50. Wang X, Luo Y, Crankshaw D, Tumanov A, Yu F, Gonzalez JE. Idk cascades: Fast deep learning by learning not to overthink. arXiv preprint arXiv:170600885. 2017.
51. Ranjan R, Patel VM, Chellappa R. Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. IEEE transactions on pattern analysis and machine intelligence. 2017;41(1):121-35.
52. Gadermayr M, Tschuchnig M. Multiple instance learning for digital pathology: A review of the state-of-the-art, limitations & future potential. Computerized Medical Imaging and Graphics. 2024:102337.
53. Shovon MSH, Mridha MF, Hasib KM, Alfarhood S, Safran M, Che D. Addressing uncertainty in imbalanced histopathology image classification of her2 breast cancer: An interpretable ensemble approach with threshold filtered single instance evaluation (sie). IEEE Access. 2023;11:122238-51.
54. Wan S, Wu T-Y, Wong WH, Lee C-Y, editors. Confnet: predict with confidence. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2018: IEEE.
55. Von Eschenbach WJ. Transparency and the black box problem: Why we do not trust AI. Philosophy & Technology. 2021;34(4):1607-22.
56. Poon AI, Sung JJ. Opening the black box of AI-Medicine. Journal of gastroenterology and hepatology. 2021;36(3):581-4.
57. Yuen KKF. A Tutorial on Explainable Image Classification for Dementia Stages Using Convolutional Neural Network and Gradient-weighted Class Activation Mapping. arXiv preprint arXiv:240810572. 2024.
58. Huff DT, Weisman AJ, Jeraj R. Interpretation and visualization techniques for deep learning models in medical imaging. Physics in Medicine & Biology. 2021;66(4):04TR1.
59. Pisula JI, Datta RR, Valdez LB, Avemarg J-R, Jung J-O, Plum P, et al. Predicting the HER2 status in oesophageal cancer from tissue microarrays using convolutional neural networks. British Journal of Cancer. 2023;128(7):1369-76.

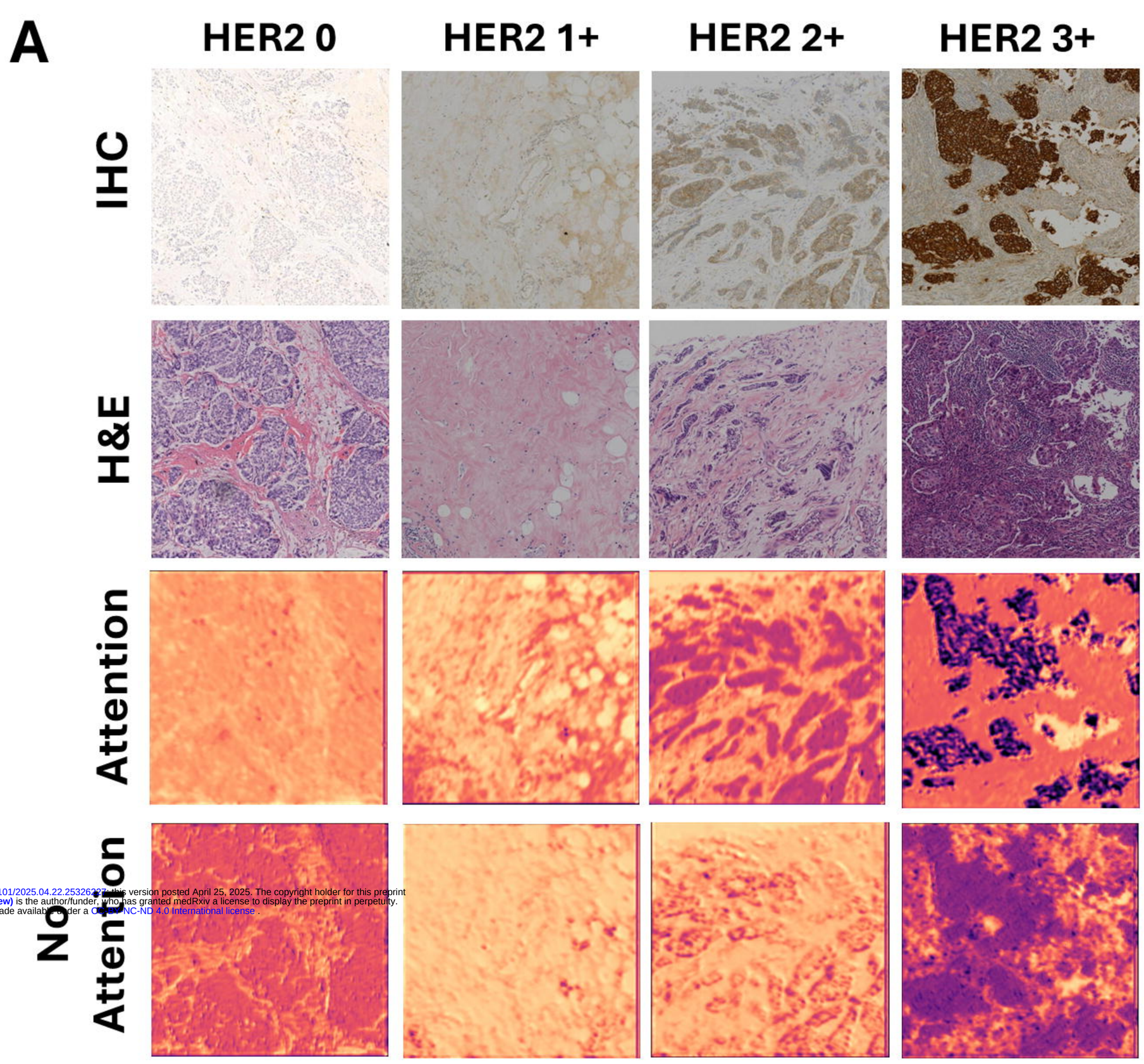
medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint
**(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

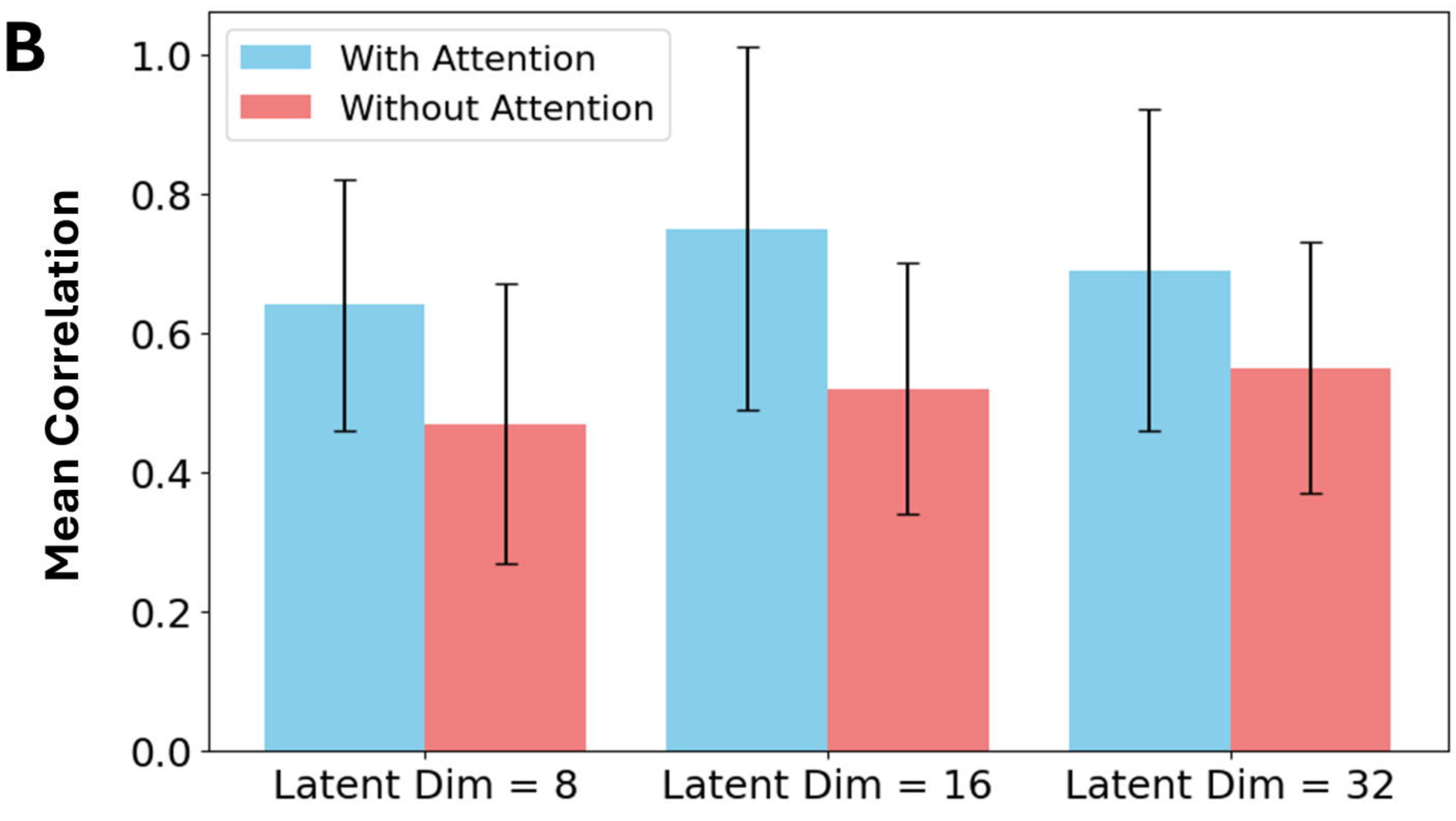
60. Jung M, Song SG, Cho SI, Shin S, Lee T, Jung W, et al. Augmented interpretation of HER2, ER, and PR in breast cancer by artificial intelligence analyzer: enhancing interobserver agreement through a reader study of 201 cases. Breast Cancer Research. 2024;26(1):31.
61. Huang Z, Zhang X, Ju Y, Zhang G, Chang W, Song H, et al. Explainable breast cancer molecular expression prediction using multi-task deep-learning based on 3D whole breast ultrasound. Insights into Imaging. 2024;15(1):227.
62. Atallah NM, Toss MS, Verrill C, Salto-Tellez M, Snead D, Rakha EA. Potential quality pitfalls of digitalized whole slide image of breast pathology in routine practice. Modern Pathology. 2022;35(7):903-10.
63. Duenweg SR, Bobholz SA, Lowman AK, Stebbins MA, Winiarz A, Nath B, et al. Whole slide imaging (WSI) scanner differences influence optical and computed properties of digitized prostate cancer histology. Journal of Pathology Informatics. 2023;14:100321.
64. Liu S, Zhu C, Xu F, Jia X, Shi Z, Jin M, editors. Bci: Breast cancer immunohistochemical image generation through pyramid pix2pix. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition; 2022.
65. Macenko M, Niethammer M, Marron JS, Borland D, Woosley JT, Guan X, et al., editors. A method for normalizing histology slides for quantitative analysis. 2009 IEEE international symposium on biomedical imaging: from nano to macro; 2009: IEEE.
66. Behar N, Shrivastava M. ResNet50-Based Effective Model for Breast Cancer Classification Using Histopathology Images. CMES-Computer Modeling in Engineering & Sciences. 2022;130(2).
67. Dutta K, Whitehead T, Laforest R, Jha A, Shoghi K. Deep learning based generation of high-count preclinical [18F]-FDG PET images from low-count [18F]-FDG PET images. Soc Nuclear Med; 2022.
68. Dutta K, Liu Z, Laforest R, Jha A, Shoghi KI, editors. Deep learning framework to synthesize high-count preclinical PET images from low-count preclinical PET images. Medical Imaging 2022: Physics of Medical Imaging; 2022: SPIE.
69. DeVries T, Taylor GW. Learning confidence for out-of-distribution detection in neural networks. arXiv preprint arXiv:180204865. 2018.



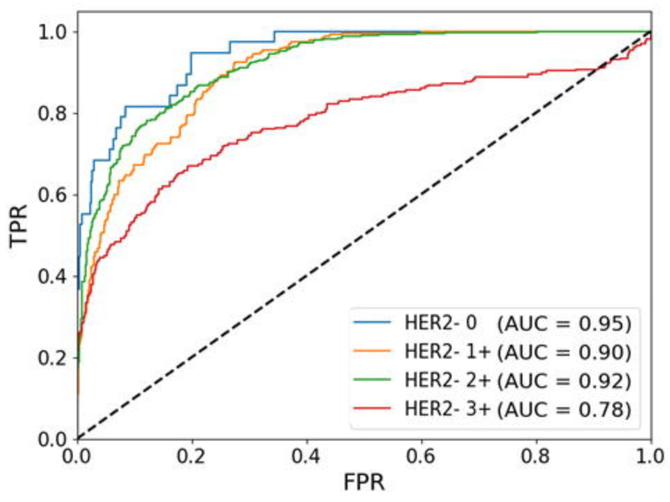
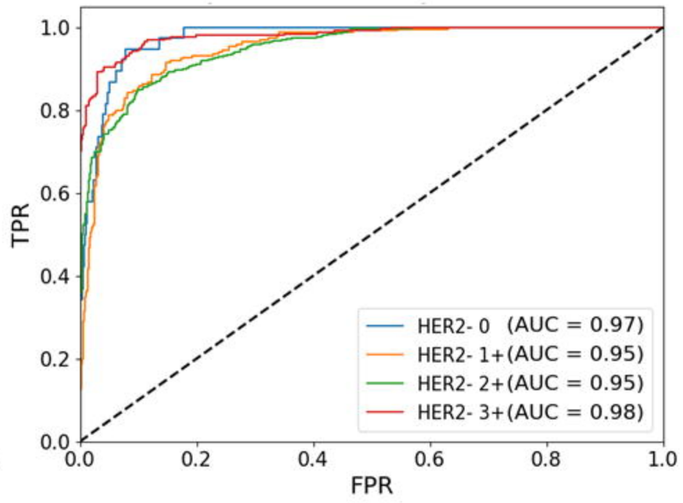
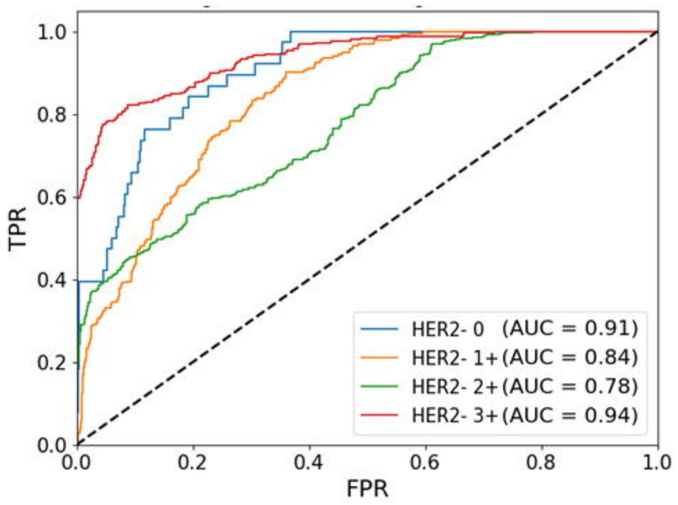
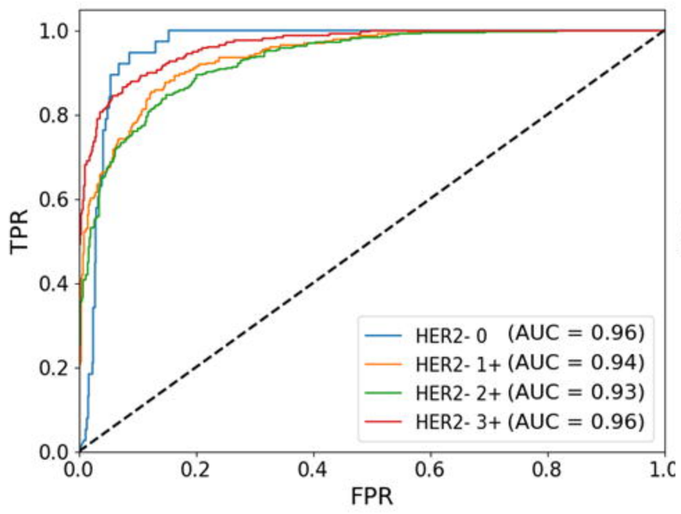
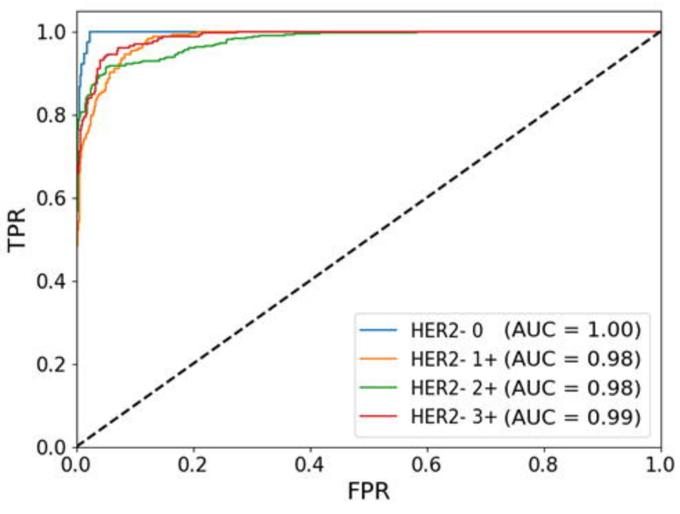
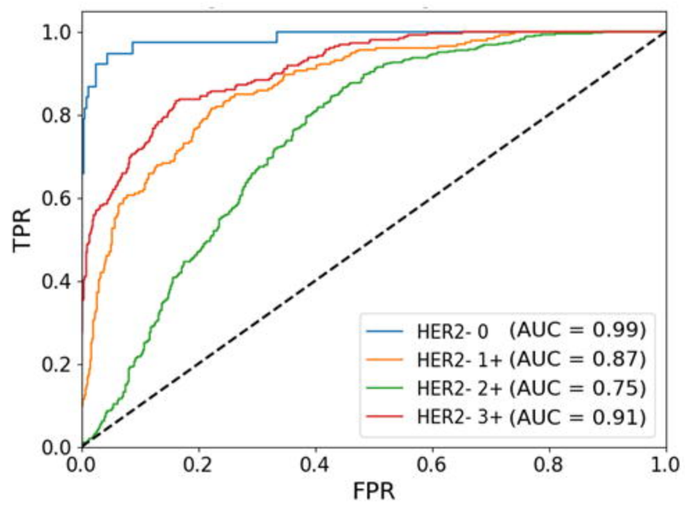
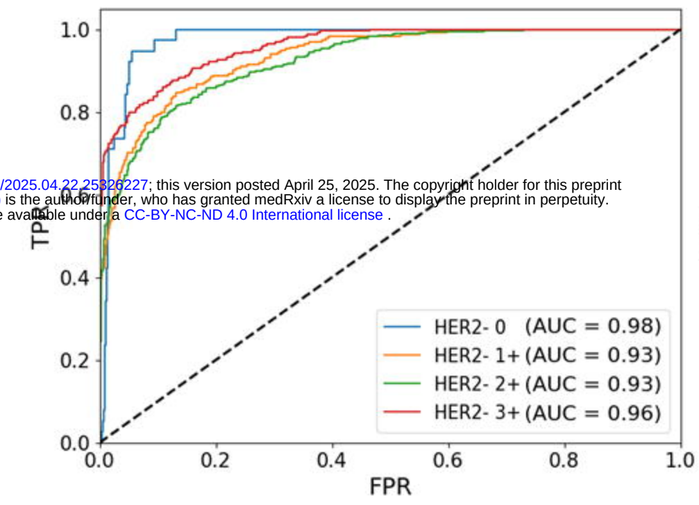
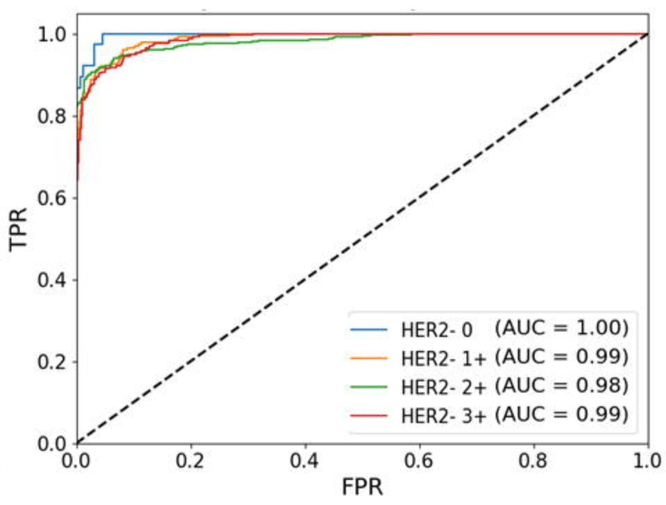
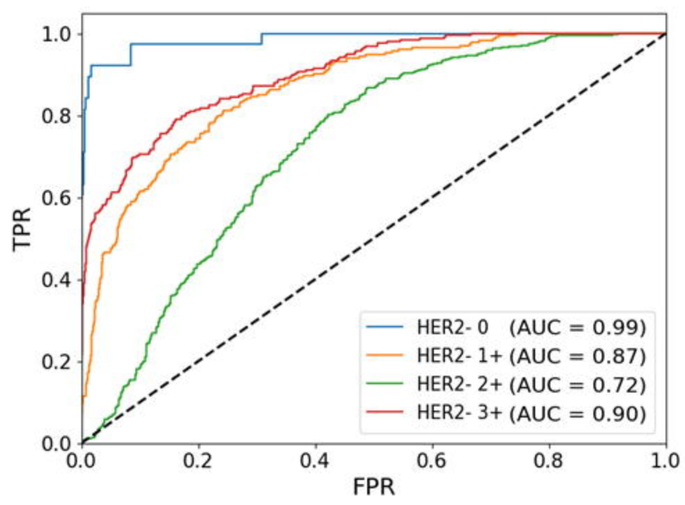
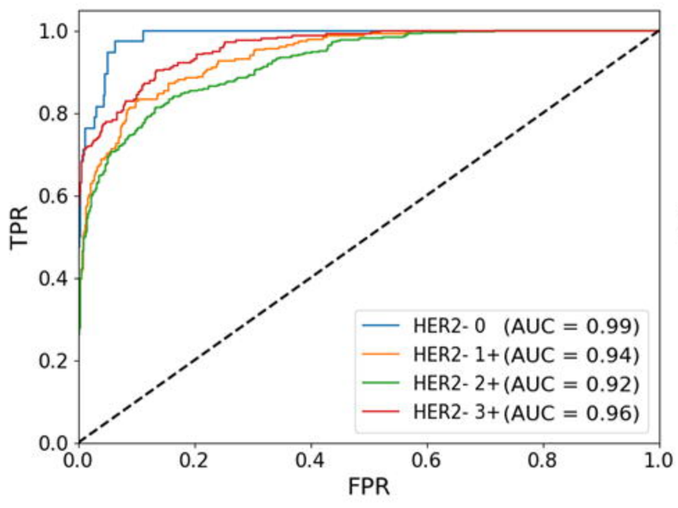
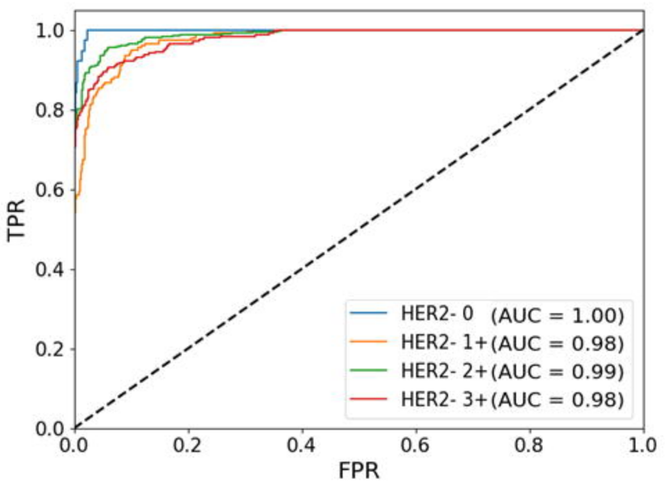
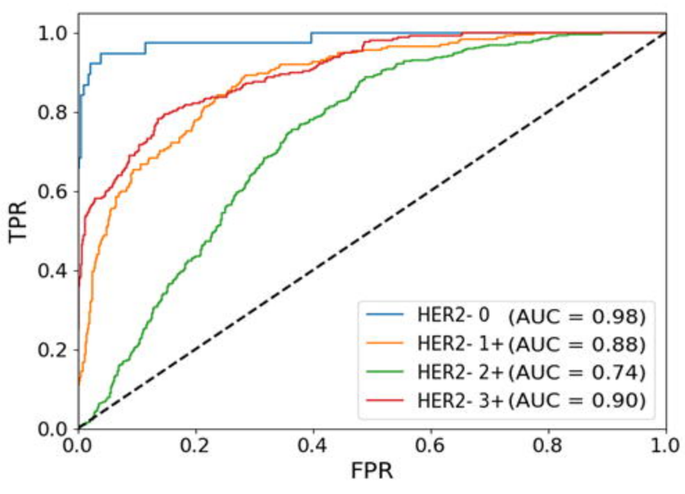
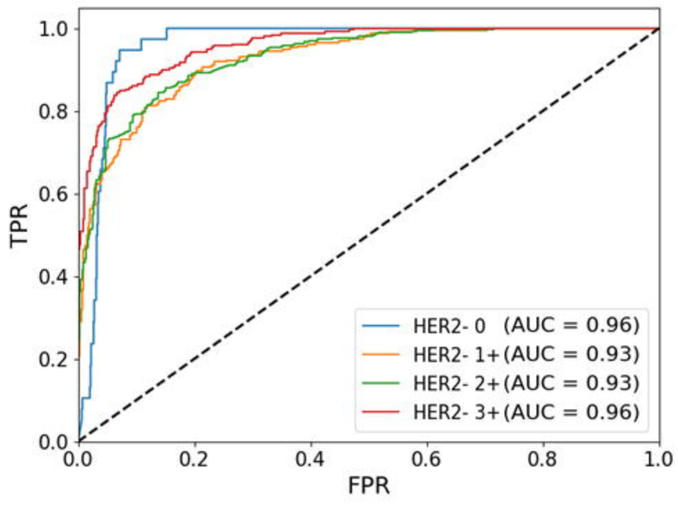
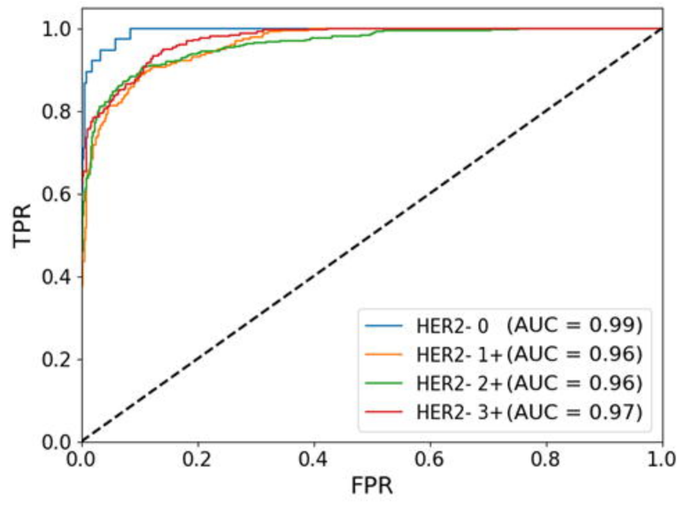
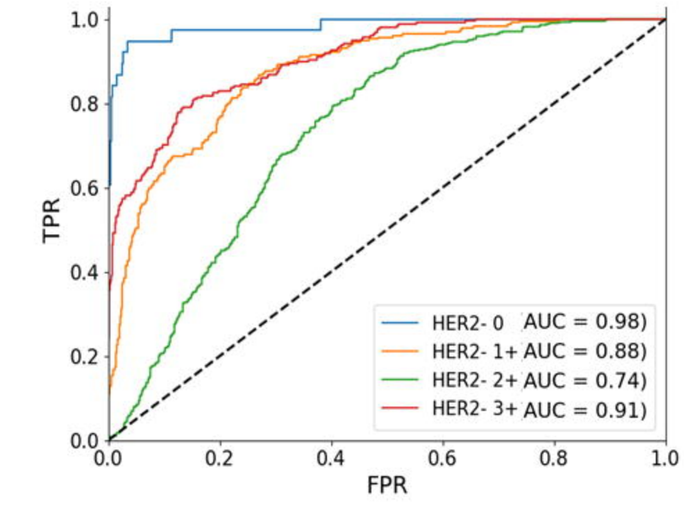

medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license

medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC BY-NC-ND 4.0 International license .

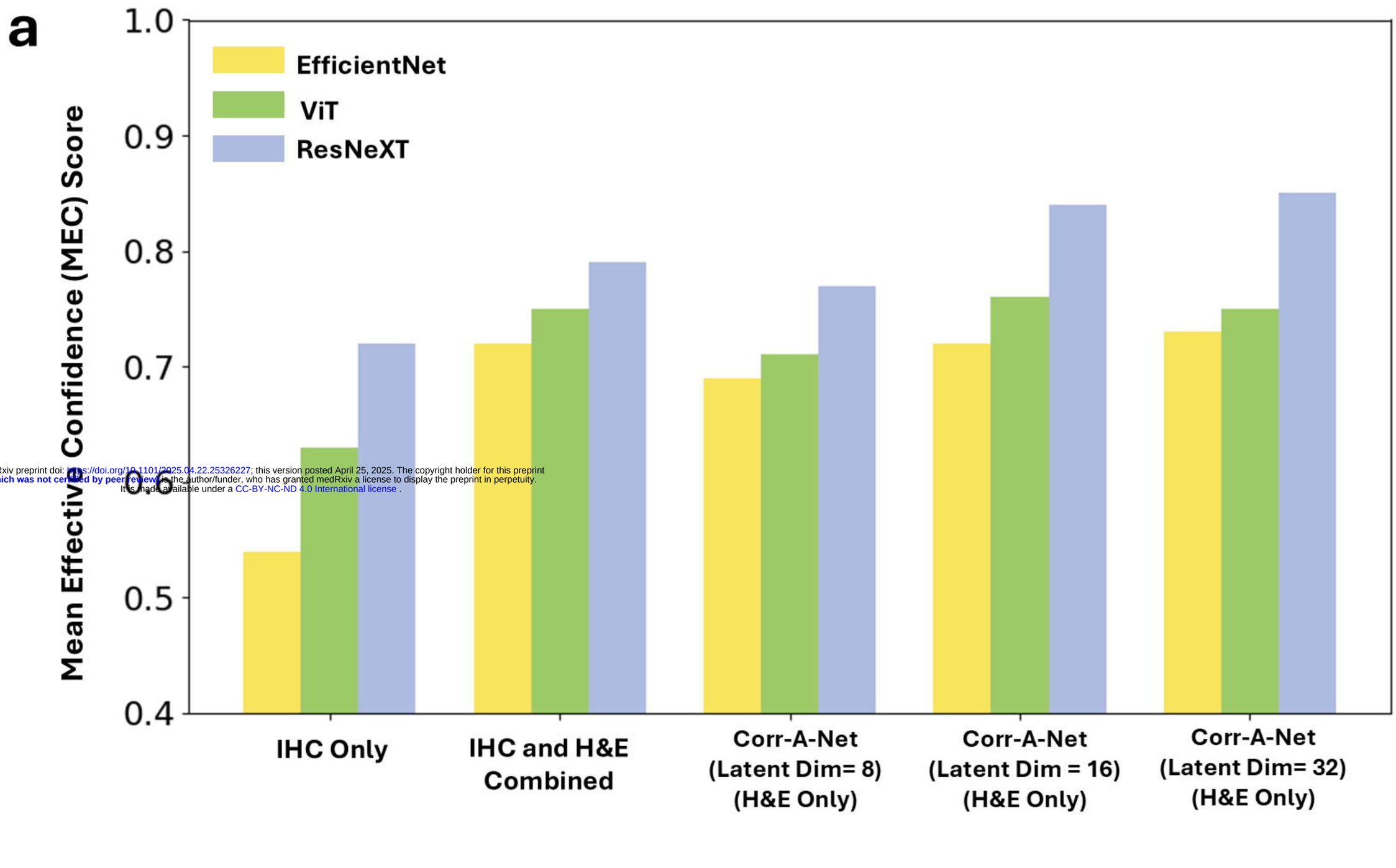


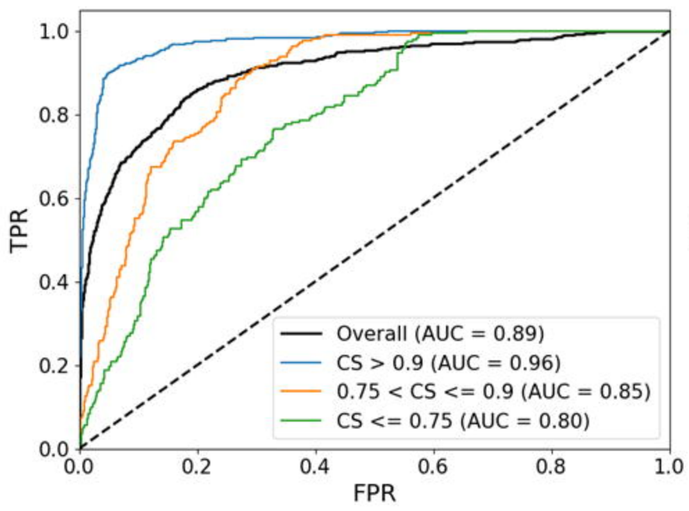
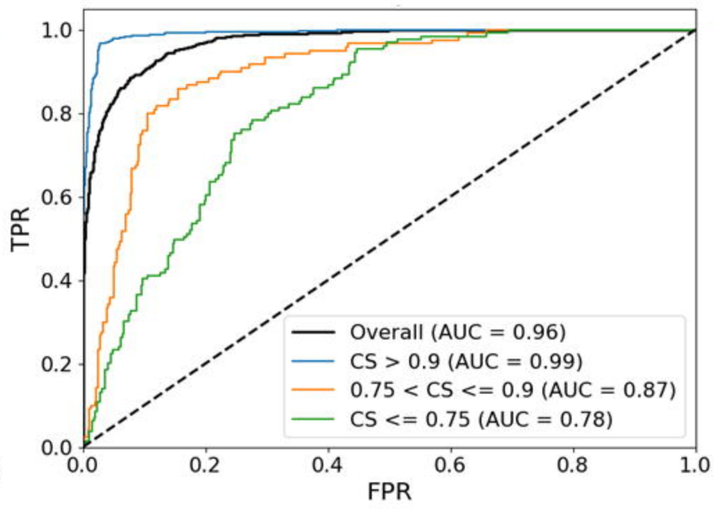
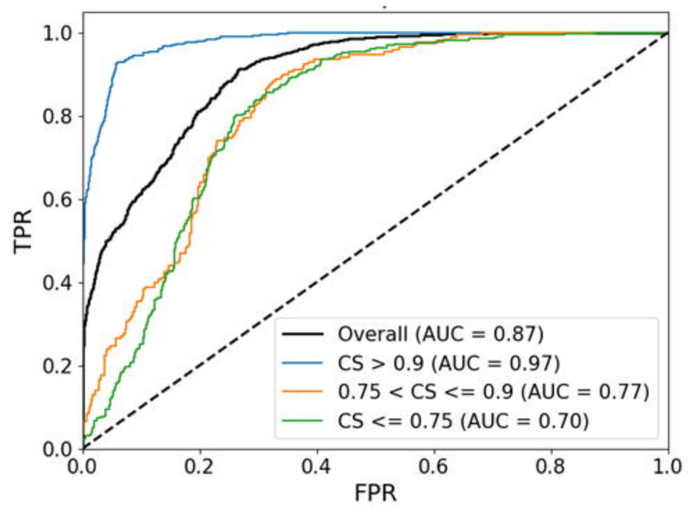
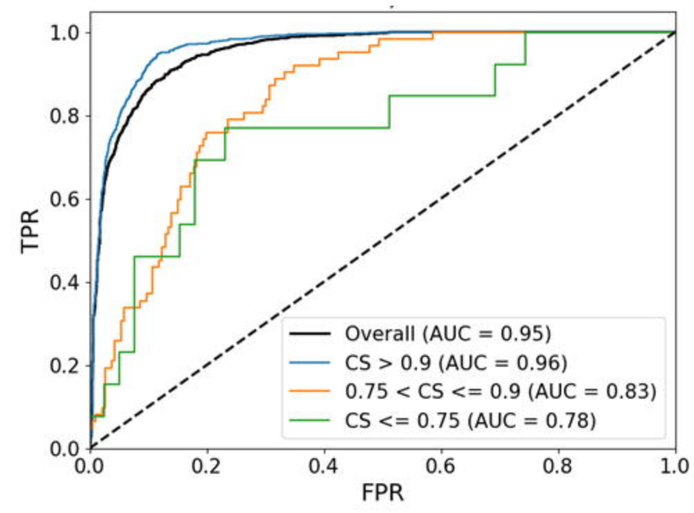
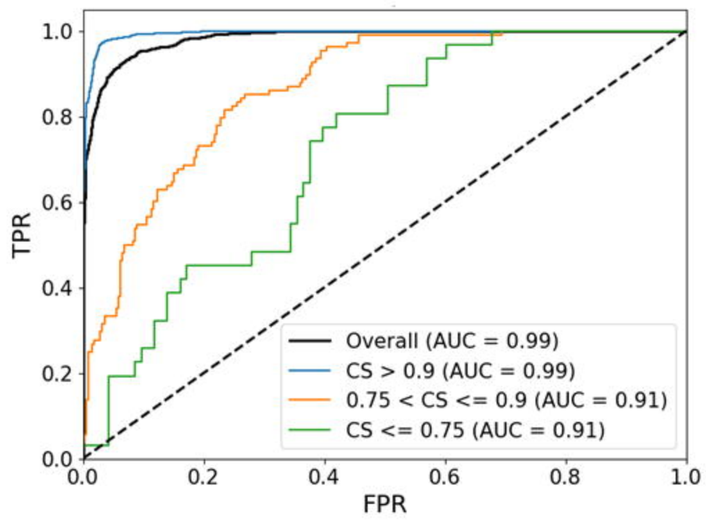
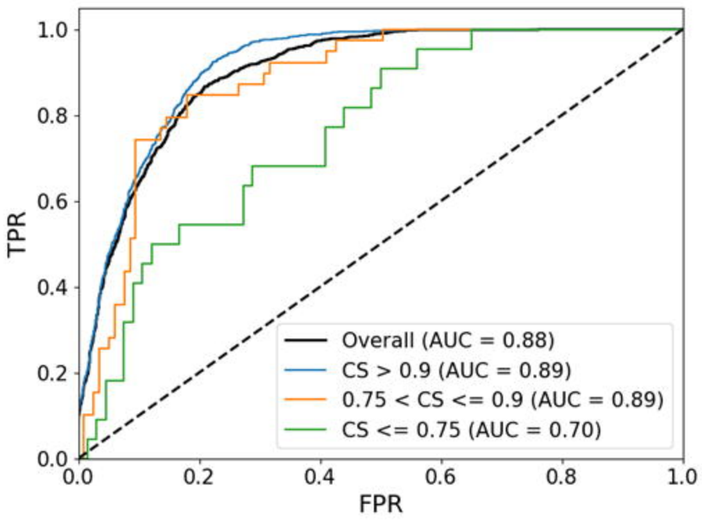
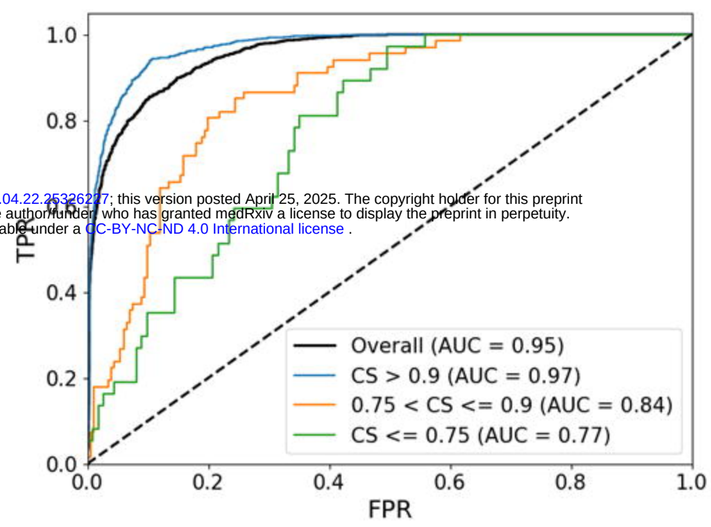
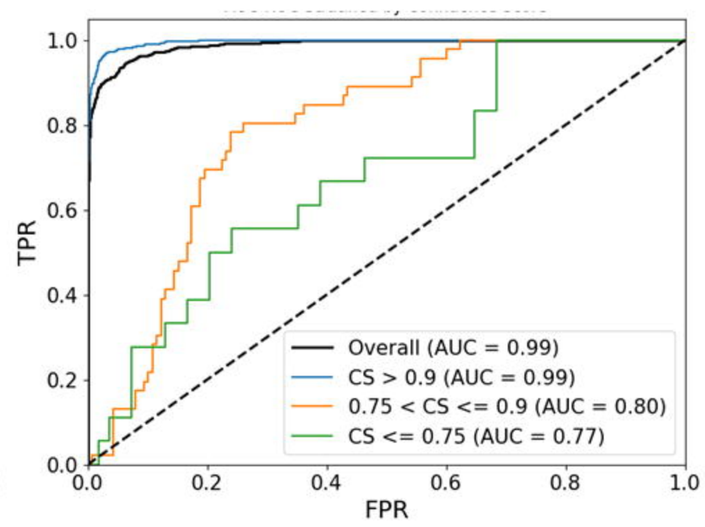
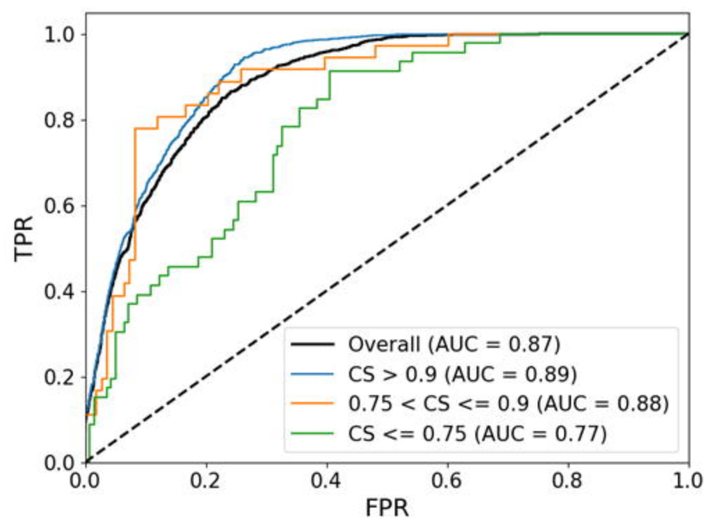
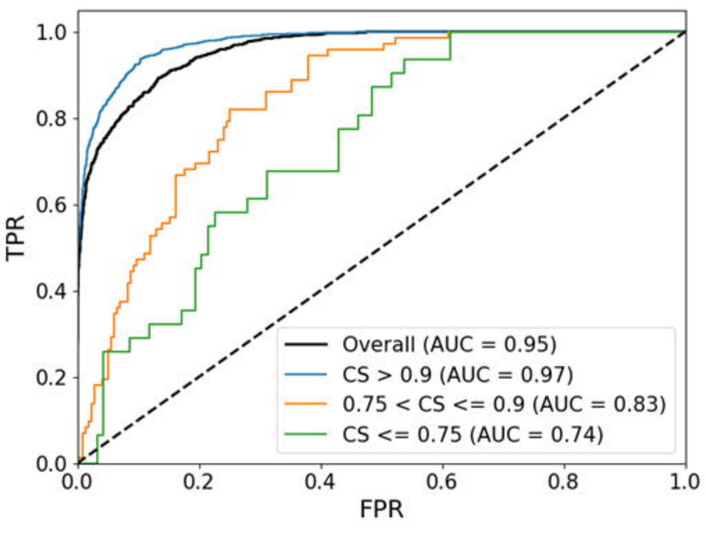
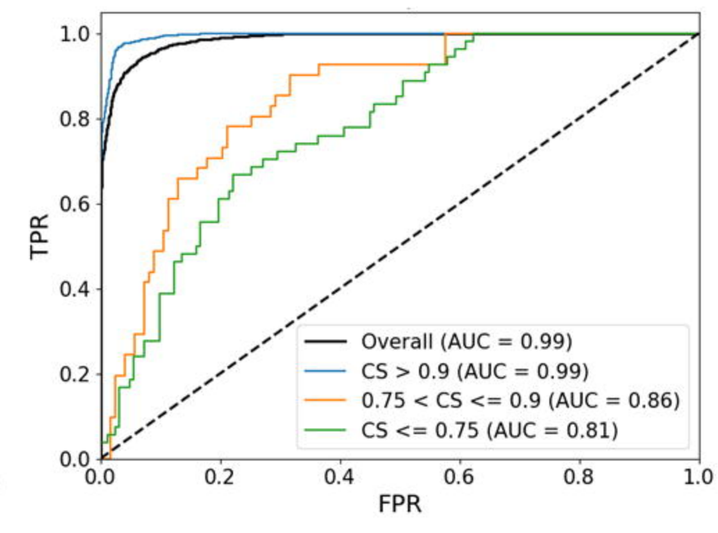
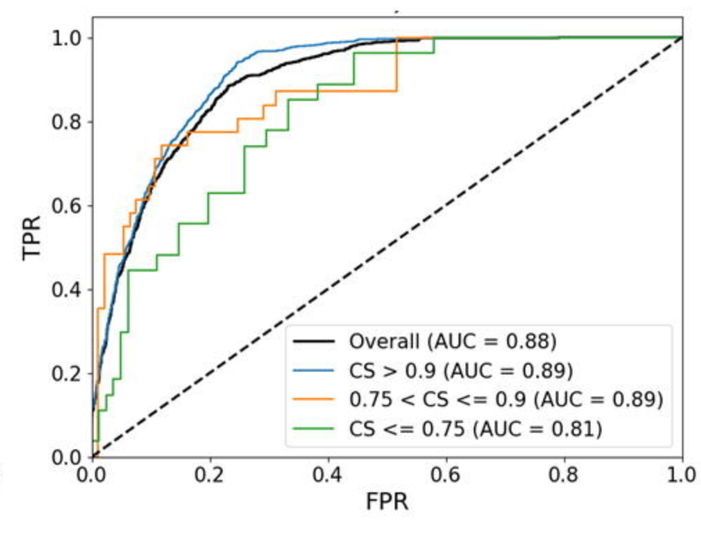
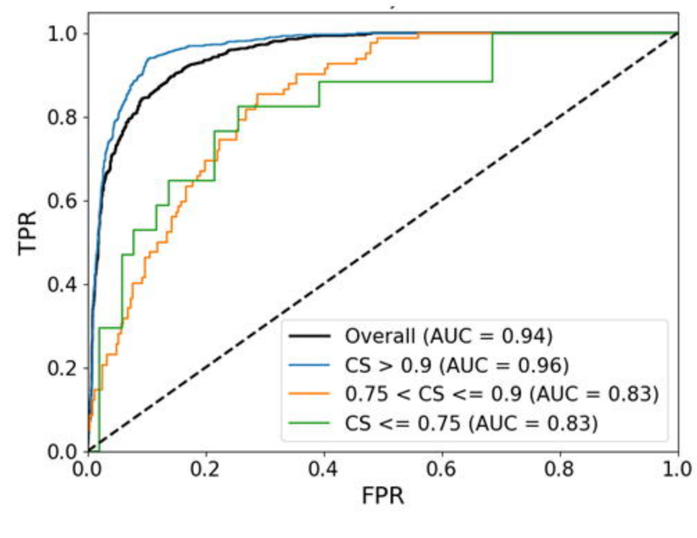
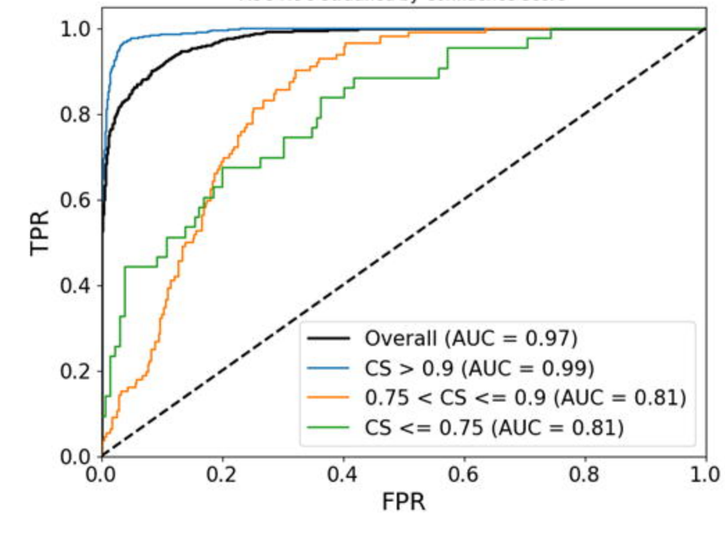
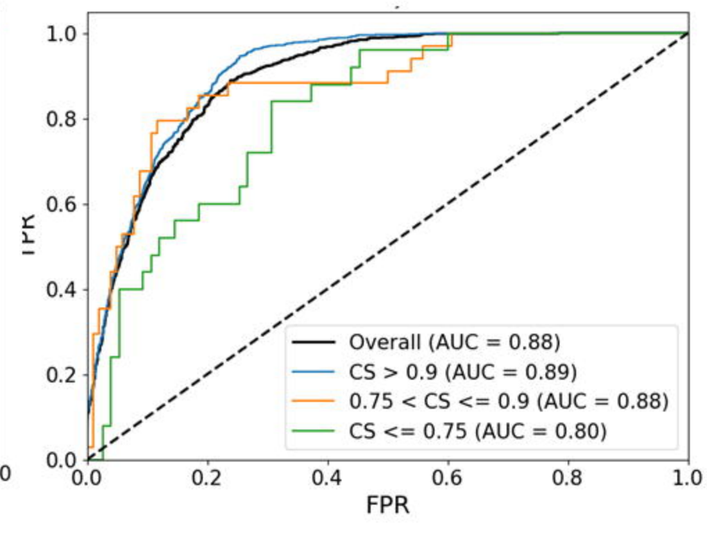
medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

| | ViT | ResNext | EfficientNet |
|---|---|---|---|
| **IHC Only** |  |  |  |
| **IHC and H&E** |  |  |  |
| **Corr-A-Net (Latent Dim = 8) (H&E Only)** |  |  |  |
| **Corr-A-Net (Latent Dim = 16) (H&E Only)** |  |  |  |
| **Corr-A-Net (Latent Dim = 32) (H&E Only)** |  |  |  |

medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .



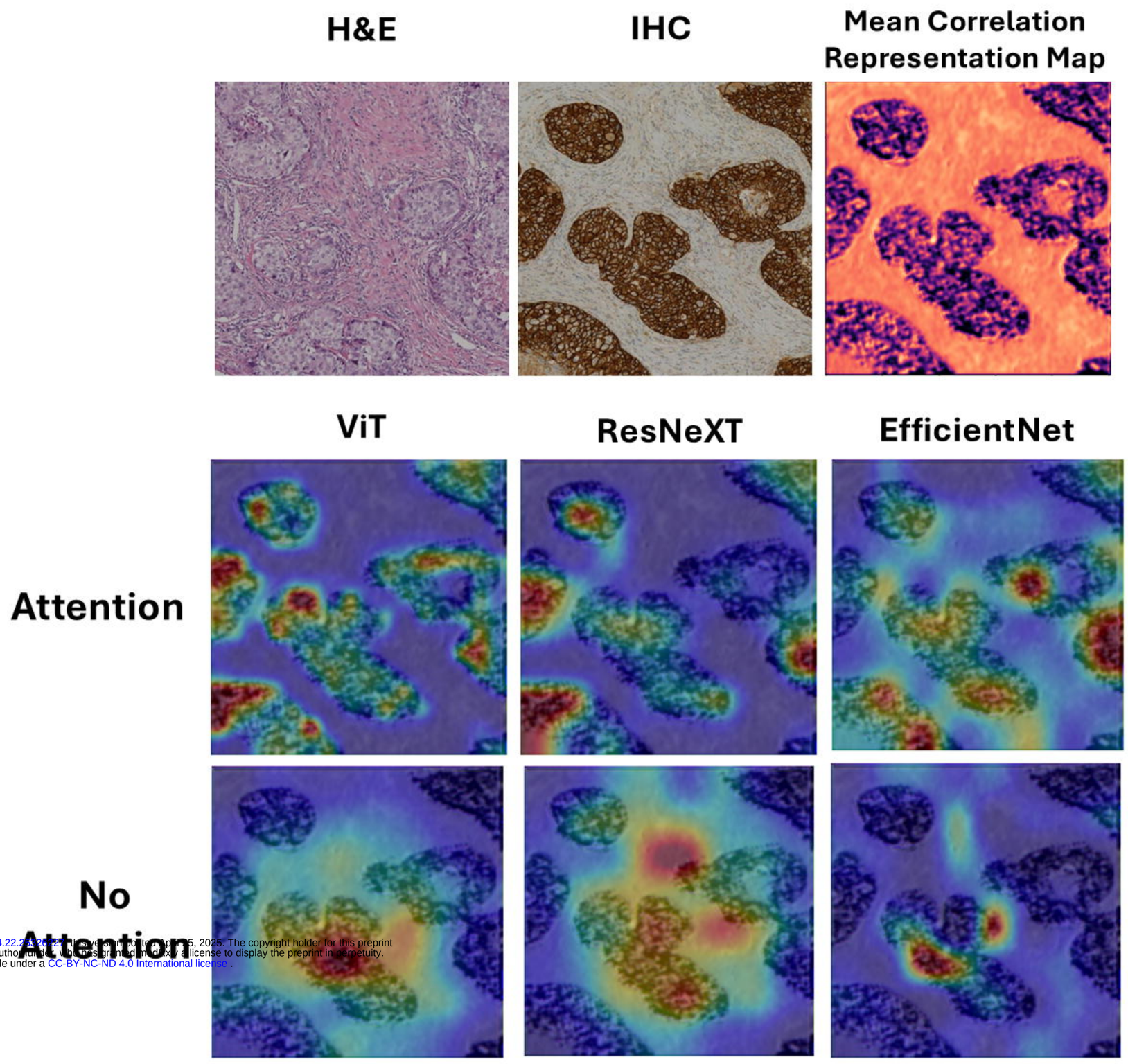
**b**

| | ViT | ResNext | EfficientNet |
|---|---|---|---|
| **IHC Only** |  |  |  |
| **IHC and H&E** |  |  |  |
| **Corr-A-Net (Latent Dim = 8) (H&E Only)** |  |  |  |
| **Corr-A-Net (Latent Dim = 16) (H&E Only)** |  |  |  |
| **Corr-A-Net (Latent Dim = 32) (H&E Only)** |  |  |  |

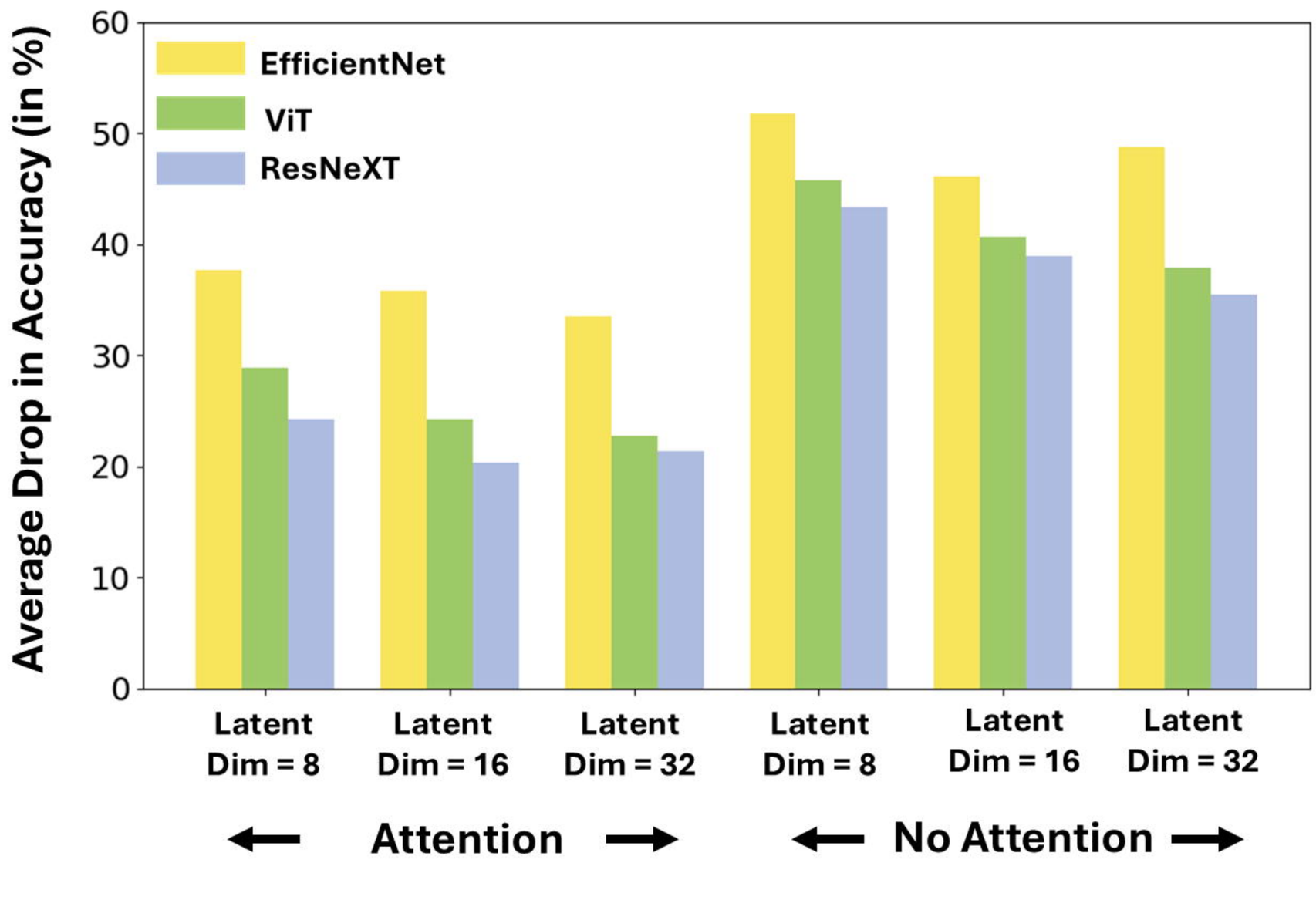
medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

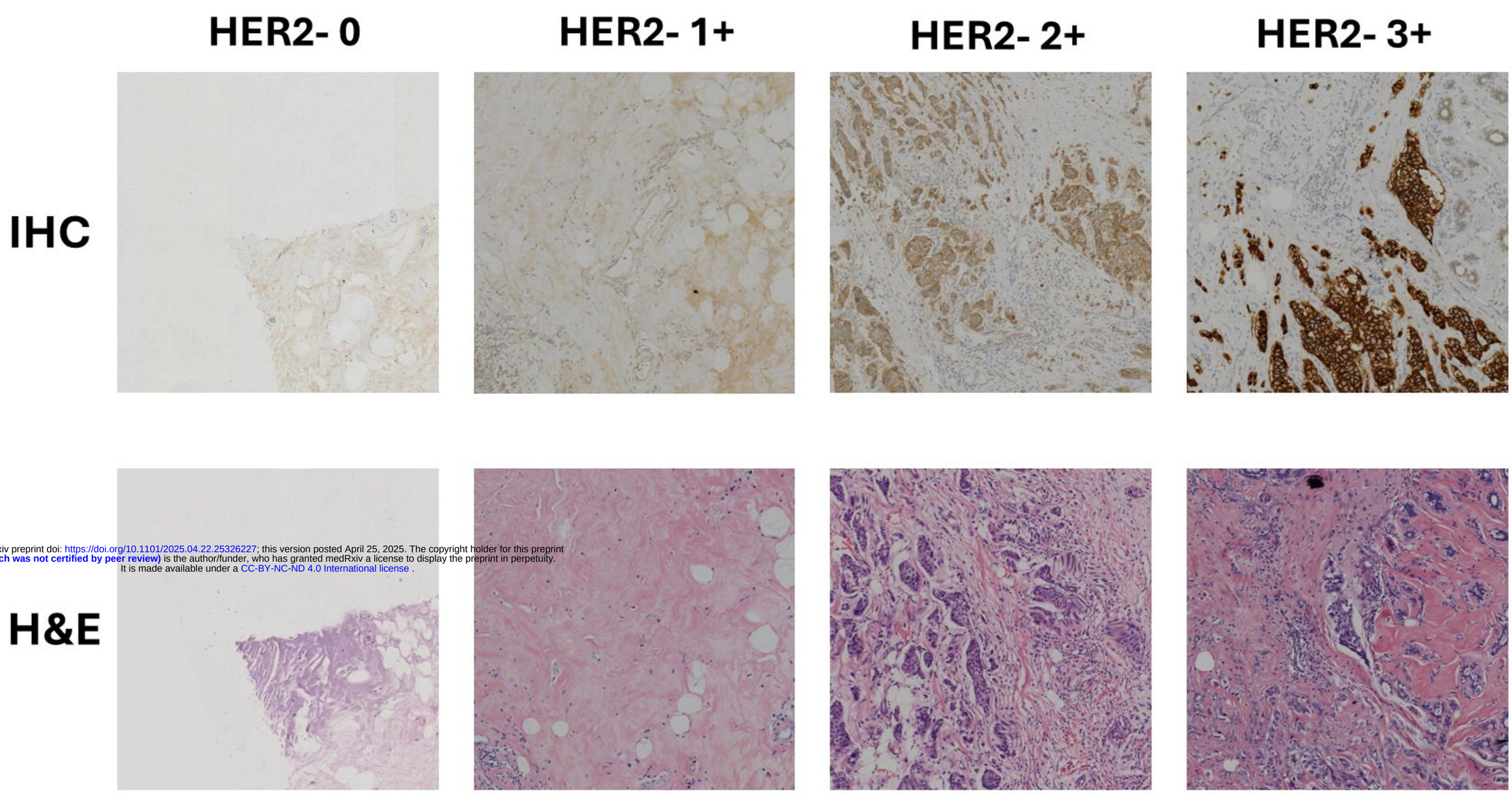
medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227 this version posted April 25, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

**A**

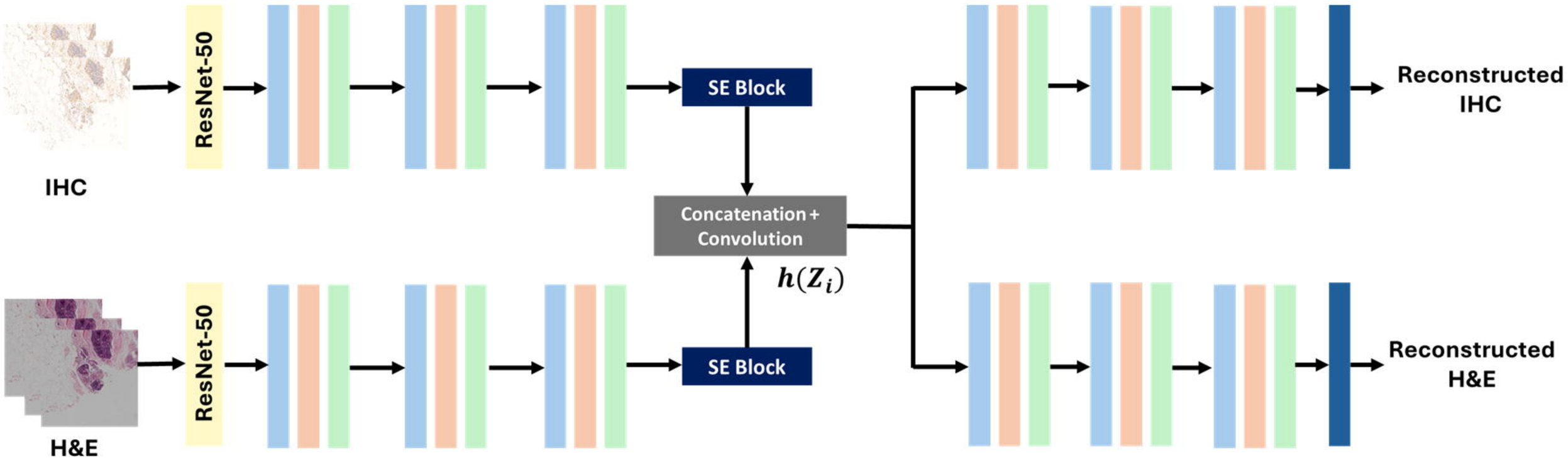



**B**






medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .




medRxiv preprint doi: https://doi.org/10.1101/2025.04.22.25326227; this version posted April 25, 2025. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .