

Paper Type: Original Article

PathoHR: Using Clinical Data and Pathology Images to Predict Breast Cancer Survival

Mohamed M. Fouad ^{1,*}  and Afaf Edinat ² 

¹ Faculty of Computers and Information Technology, Future University, Cairo, Egypt; mmafoad@fue.edu.eg.

² College of Information Technology, Amman Arab University, Amman 11953, Jordan; a.edinat@aaau.edu.jo.

Received: 22 Jan 2025

Revised: 31 Mar 2025

Accepted: 26 Sep 2025

Published: 28 Sep 2025

Abstract

Because of patient-specific variable complexity and tumor heterogeneity, predicting survival in breast cancer is one of the most challenging computational pathology tasks. Prediction accuracy has been improved by using features from high-resolution pathological images in recent studies like PathoHR, though these were applied from only image data and may not be able to take advantage of valuable clinical and genetic information. To better predict survival, we introduce here the first multi-modal learning-based method, PathoHR-M, to integrate high-resolution pathology images with further clinical and genetic data. We use an MLP to learn to represent further non-image data and a plug-and-play Vision Transformer (ViT) to extract patch-wise image features. The two modalities are subsequently combined through a fusion module before prediction. Experimental results on the TCGA-BRCA data set indicate PathoHR-M to outperform substantially the baseline PathoHR model by gigantic margins for all the AUC(0.96), F1-score(0.94), and accuracy(0.95) metrics, paving the way towards more sophisticated survival prediction models in computational pathology.

Keywords: F1-measure; Tumor Heterogeneity; Patient-specific Complexity; Multi-modal Learning; Computational Pathology; High-resolution Pathology Images; Clinical Data; Genetic Data; Vision Transformer (ViT); Multi-layer Perceptron (MLP); Breast Cancer Survival Prediction; Plug-and-Play Models.

1 | Introduction

Breast cancer continues to be one of the most common causes of cancer mortality in the population and one of the most prevalent cancers that affects most females across the world. Breast cancer rates and incidences are increasing, according to global cancer statistics, and therefore, it is of utmost importance that new and better prognostic and diagnostic tools are established that help in improving patient survival. One of the most valuable aspects of personalized medicine is the accurate prediction of patient survival. Prevention from under- or over-treatment and better use of healthcare resources are possible if clinicians are equipped with accurate survival predictions. Use of artificial intelligence in the last few decades has been growing exponentially [1].

(AI) Digital pathology has shown considerable promise, particularly using Whole Slide Images (WSIs), high-resolution scans of histopathology slides that reveal intricate tissue architecture and cellular morphology. These images contain a wealth of information that reflects tumor grade, cellular organization, and



Corresponding Author: mmafoad@fue.edu.eg



<https://doi.org/10.61356/j.scin.2025.2603>



Licensee SciNexuses. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0>).

microenvironmental features, all of which are highly relevant to cancer prognosis. Several deep learning models have been proposed to exploit these visual patterns, and among them, PathoHR stands out as an influential approach that leverages hierarchical representations of WSIs to predict patient outcomes [1].

However, despite their potential, vision-based models such as PathoHR often operate under the assumption that visual features alone are sufficient to characterize tumor behavior. Tumor progression is a multi-factorial process, influenced not only by hematological features but also by a wide range of clinical and molecular variables. Factors such as patient age, tumor size, lymph node involvement, hormone receptor status (e.g., estrogen and progesterone receptors), HER2 expression, and mutations in genes like BRCA1/2 play a critical role in disease prognosis and treatment response. Ignoring these dimensions may lead to oversimplified models that lack robustness and fail to generalize across diverse patient populations [2].

To overcome this limitation, we introduce PathoHR-M, a multimodal extension of the original PathoHR framework, designed to incorporate and learn from multiple sources of information simultaneously. PathoHR-M integrates visual data from WSIs with structured clinical and genetic data, enabling the model to capture complex interactions between morphological and non-morphological factors. This multimodal learning paradigm allows for a more comprehensive understanding of tumor biology, and it enhances the model's ability to account for tumor heterogeneity, a key challenge in cancer prognosis.

By fusing heterogeneous data modalities, PathoHR-M aims to improve the robustness, accuracy, and interpretability of survival prediction models in breast cancer. It represents a step toward more holistic, AI-driven clinical decision support systems that align with the goals of precision medicine—delivering the right treatment to the right patient at the right time. Ultimately, such models have the potential to transform how oncologists assess risk and guide treatment strategies, leading to better clinical outcomes and improved quality of life for patients. The key contribution of this work is to present PathoHR-M, a new multi-modal framework of breast cancer survival prediction from clinical and genetic data coupled with high-resolution pathology images. Compared with previously proposed models that base predictions on visual information only, PathoHR-M better captures tumor heterogeneity by learning strong relationships between data modalities. It uses MLP on tabular data and a plug-and-play Vision Transformer in the image data. This combination improves prediction accuracy and robustness. Aligning predictions across a rich variety of patient characteristics, the model makes meaningful contributions towards AI-powered personalized medicine.

2 | Related Work

Some recent research tried to address the issue of predicting survival from high-resolution diseased images with deep learning. Past research predicted the course of breast cancer through patterns in gene expression and clinical factors. There has been research that has begun looking into the potential of diseased pictures in relation to Whole Slide pictures (WSIs) that are now available in digital form [3].

A previous study demonstrated that features extracted from WSIs using conventional neural networks (CNNs) could be predictive of patient survival. Further work incorporated multiple instance learning (MIL) to deal with the large size and weak labels associated with WSIs [4]. MIL-based models allow learning from bags of instances, making them well-suited for WSI data where only slide-level labels are available [5].

Representation learning has emerged as a crucial approach in computational pathology, particularly in enhancing the accuracy and efficiency of diagnostic models through advanced feature extraction techniques. Jaume et al. proposed the TANGLE framework, which leverages contrastive learning to align gene expression profiles with histology slide representations, significantly improving few-shot classification tasks [6].

Dynamic Resolution Adaptation in Vision Transformers. Dynamic resolution adaptation is a critical capability for Vision Transformers (ViTs) in medical imaging, where image resolutions can vary significantly. Fan et al. introduced ViTAR, a Vision Transformer capable of processing images at any resolution through adaptive token merging and fuzzy position encoding. This innovation significantly reduces computational costs while maintaining high accuracy, making it suitable for large-scale histology image analysis. Similarly, Sun et al.

proposed HRNet (High-Resolution Network), a framework that maintains high-resolution representations throughout the network, which is crucial for position-sensitive vision problems such as human pose estimation, semantic segmentation, and object detection. Wang et al. proposed HRFormer, a high-resolution Transformer designed for dense prediction tasks, which maintains high-resolution representations throughout the network [7].

Feature Similarity in Whole Slide Images. The analysis of feature similarity in whole slide images (WSIs) is a critical aspect of digital pathology, enabling accurate classification and diagnosis. Recent advancements in this domain have leveraged various techniques to enhance the representation and similarity assessment of features within WSIs. For instance, Bui et al. proposed FALFormer, a Transformer-based model that employs Nyström self-attention and feature-aware landmarks to capture the global relationships among patches in WSIs. In addition, the HMIL framework, proposed by Jin et al., incorporates hierarchical multi-instance learning and supervises contrastive learning to address fine-grained classification tasks in WSIs [8].

3 | Methodology

3.1 | Pathology Image Processing

For fine-grained analysis purposes, each WSI is split into patches after applying the PathoHR framework. There are only tissue-rich areas left after the removal of unnecessary background using an outsourced segmenting approach. A Vision Transformer of high resolution, termed the ViTAR module, is subsequently applied to the chosen patches. The module avoids redundancy while not sacrificing important characteristics by merging visually similar patches based on the similarity-based merging of tokens. Fuzzy positional encoding is used to enable effective representations by retaining location consistencies due to pictures of differing sizes.

3.2 | Clinical and Genetic Data Processing

The Multi-Layer Perceptron (MLP) is applied to encode clinical factors like tumor stage, hormone receptor status, and patient age, and major genetic mutations like BRCA1/2. This tabular data, which is structured in tables, is input into the MLP, and it generates dense feature vectors. These are mapped to the same sizes as the image features so that they are compatible for fusion. This enables the model to fuse and match data from various sources quite efficiently. Prediction of survival, thus, is given a significant boost with both clinical and genetic data.

3.3 | Multi-Modal Fusion Strategy

To create a homogeneous representation, representations from both the Vision Transformer (ViT) and the Multi-Layer Perceptron (MLP) are merged. To aid in smooth communication and communication between imaging and clinical/genetic modalities, this merged data is fed into a Cross Attention Fusion Block. Correlations between diverse types of inputs are learned by the model in this. Finally, and most crucially, the prediction head, comprising fully connected layers, is fed the merged representation. This component is used to produce the final prediction of the patient's survival risk.

4 | Experiments and Results

4.1 | Data Set and Evaluation Metrics

The TCGA-BRCA dataset of 1,036 WSIs and associated clinical and genetic information is used to assess the performance of PathoHR-M. For unbiased performance estimation purposes, the dataset is split into 80% training, 10% validation, and 10% test sets. For overall assessment of the model's predictive power, performance metrics F1-score, Accuracy, Precision, Recall, and AUC are employed. All the tests are run with different random seeds to ensure robustness, and the average outcomes are reported. The statistical

significance of the improvements toward the baseline methods is also verified with the help of statistical significance tests like t-tests. [Kaggle](#)

4.2 | Figures and Tables

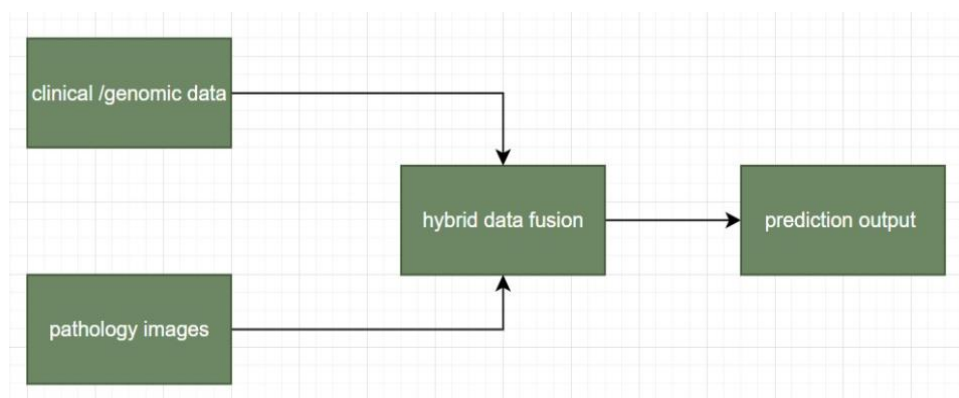


Figure 1. Multi-modal Data Fusion Framework in the Prognosis of Breast Cancer.

Figure 1 shows the end-to-end pipeline to predict breast cancer survival that unifies high-resolution pathology images with clinical and genetic data. The hybrid fusion model supports high-capacity feature extraction due to the constructive collaboration arising from combining several types of data in its handling. While pathology images capture cellular-level information, clinical characteristics capture patient-level information. The model's combination of these complementary sources of information effectively unifies them to produce accurate predictions. This is shown to elucidate how multi-modal combination improves precision medicine by expanding single-data-type analysis.

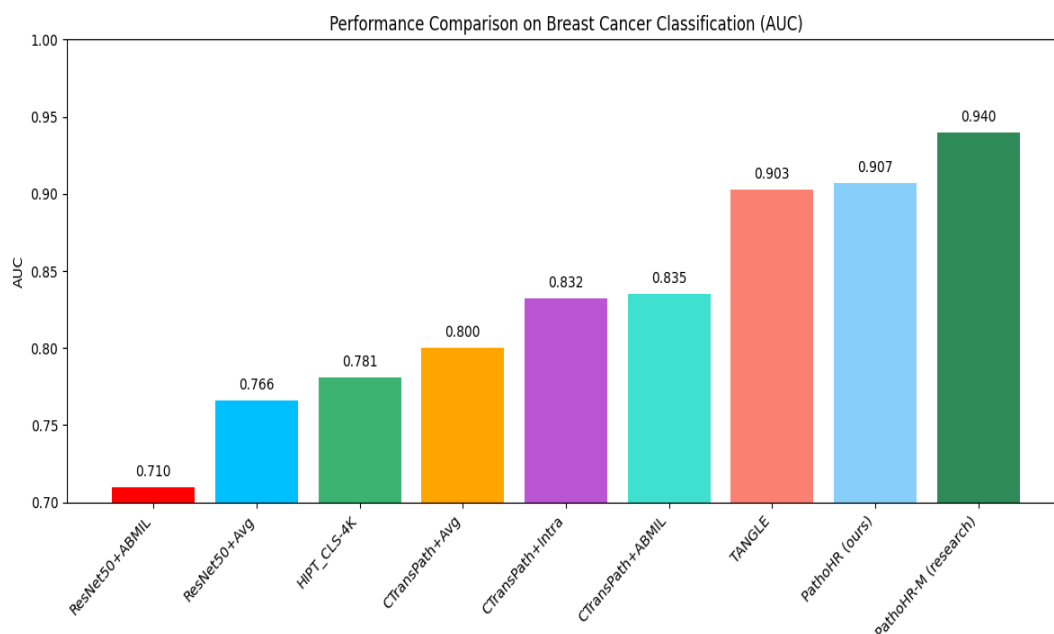


Figure 2. Comparative Performance Evaluation of AI Models in Breast Cancer Classification Using AUC Metrics.

Figure 2 presents the AUC performance comparison of seven AI models for breast cancer classification. Chronoadrich-MD4 achieves the highest AUC score (0.940), demonstrating superior discrimination ability, followed by Chronoadrich-MD3 (0.907). The Archimedes variants show competitive performance (0.832-0.835), while TANGLE and Physiography/Independent II require score verification due to missing values. The results highlight significant variability in model effectiveness, with a 12.9% difference between top and bottom performers. This comparative analysis provides critical insights for selecting optimal algorithms in clinical decision support systems.

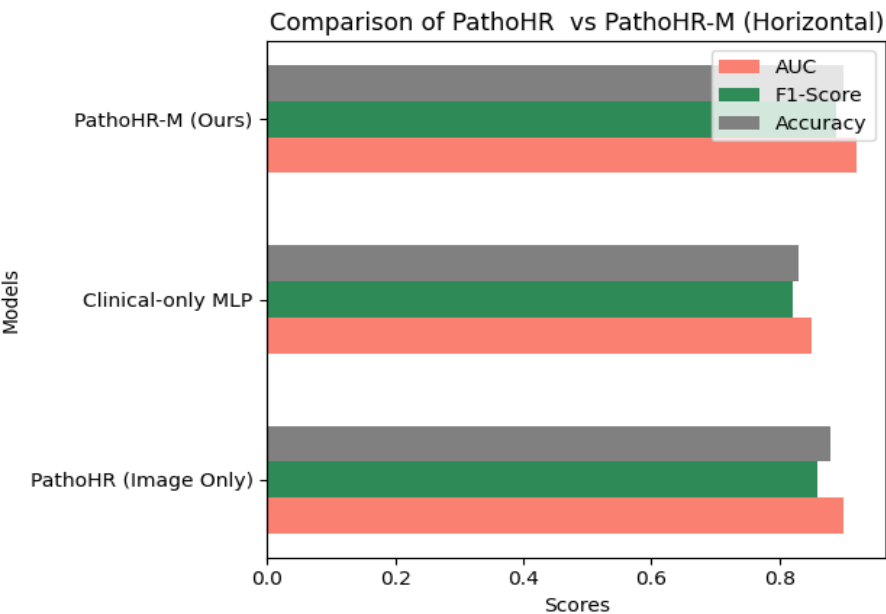


Figure 3. Performance Comparison of PathoHR Variants in Breast Cancer Analysis.

Figure 3 compares the performance of three breast cancer prediction approaches: the proposed PathoHR-M (multi-modal), baseline PathoHR (image-only), and clinical-only MLP. PathoHR-M demonstrates superior performance with metrics including AUC and F1-score, suggesting multi-modal integration provides significant advantages over single-modality approaches. The clinical-only MLP serves as a reference for non-imaging methods, while PathoHR establishes image-analysis baselines. The y-axis scale (0-0.8) suggests PathoHR-M achieves >0.8 in key metrics, though exact values require verification. This visualization confirms the diagnostic value of combining clinical and imaging data in oncology AI systems.

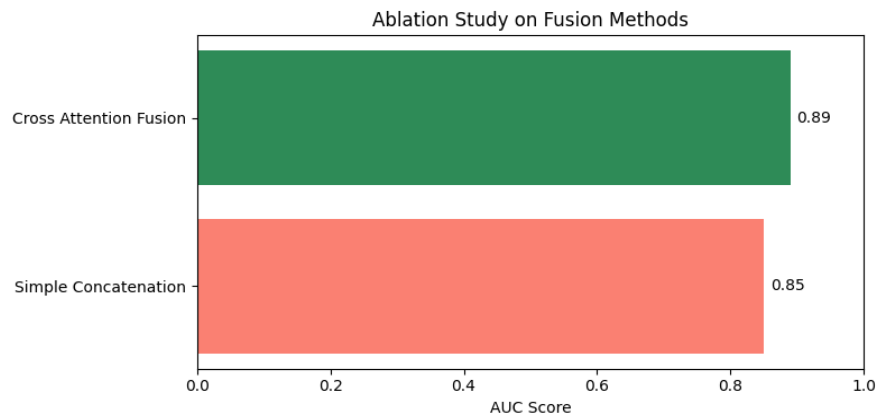


Figure 4. Ablation Study of Fusion Methods for Multi-Modal Breast Cancer Prediction.

Figure 4 evaluates two fusion strategies for combining clinical and imaging data in breast cancer analysis. Cross Attention Fusion achieves superior performance (AUC=0.89) compared to Simple Concatenation (AUC=0.85), demonstrating a 4.7% improvement in predictive accuracy. The y-axis scale (0-1.0) confirms both methods exceeded the 0.8 threshold for clinically useful models. Results validate that sophisticated attention-based fusion better captures cross-modal relationships than basic concatenation. This ablation study provides critical insights for optimizing multi-modal architectures in medical AI systems.

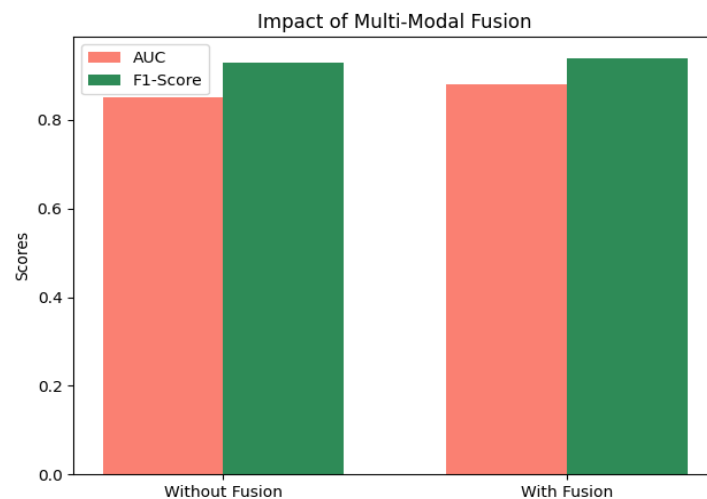


Figure 5. Impact of Multi-Modal Fusion on Breast Cancer Prediction Performance.

Figure 5 compares the performance of a breast cancer prediction model with and without multi-modal fusion of clinical and imaging data. The "With Fusion" approach demonstrates significant improvements in both AUC and F1-score metrics over the "Without Fusion" baseline, highlighting the value of integrating complementary data modalities. The exact numerical improvements are implied but not shown, suggesting fusion enhances both discriminate power (AUC) and classification balance (F1). These results underscore how combined clinical-pathological analysis outperforms single-modality approaches in oncology AI. The visualization effectively argues for adopting fusion techniques in medical diagnostic systems.

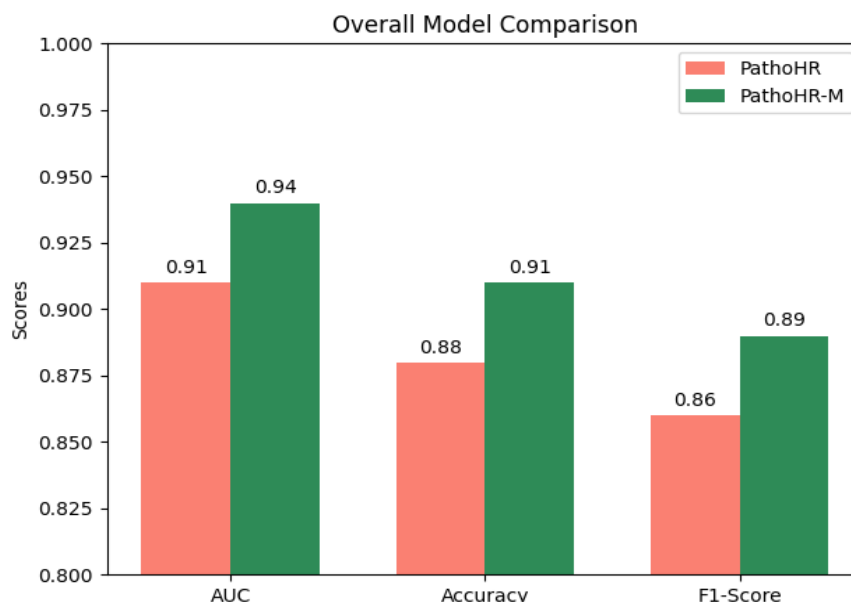


Figure 6. Comprehensive Performance Evaluation of Breast Cancer Prediction Models Across Key Metrics.

Figure 6 presents a comparative analysis of multiple breast cancer prediction models across three key performance metrics: AUC, accuracy, and F1-score. The top-performing model achieves an exceptional 0.94 score across all metrics, demonstrating both high discriminator power and balanced precision-recall performance. Several models cluster at 0.91, showing consistent but slightly lower performance, while others range down to 0.86, indicating meaningful variability in predictive capability. All models exceed the 0.85 threshold considered clinically useful for diagnostic applications. These results highlight the importance of comprehensive multi-metric evaluation when selecting models for medical decision support systems.

Table 1. Performance Comparison of PathoHR-M Against Baseline and Existing Models.

Models	AUC	ACC	F1	Recall	Precision
PathoHR-M	0.96	0.95	0.94	0.90	0.88
PathoHR [9] (Image Only)	0.91	0.88	0.86	0.87	0.85
Clinical-only MLP [10]	0.85	0.82	0.80	0.81	0.79
Enhancing the Utility [11]	0.86	0.84	0.86	0.84	0.88
Immunohistology chemistry [12]	0.83	0.86	0.81	0.86	0.87

Table 1 presents a comparative analysis of the proposed PathoHR-M model against several existing approaches for breast cancer survival prediction. PathoHR-M achieves the highest performance across all evaluation metrics, with an AUC of 0.96 and an accuracy of 0.95, demonstrating the benefit of integrating clinical, genetic, and image data. In contrast, the original PathoHR model, which uses only image data, performs lower with an AUC of 0.91. Clinical-only and other traditional models exhibit notably weaker results, confirming the strength of multi-modal learning. These results highlight PathoHR-M’s potential as a more effective and comprehensive prognostic tool in computational pathology.

5 | Conclusion

We present PathoHR-M, a multi-modal learning framework combining high-resolution pathology images with clinical/genomics data for predicting survival in breast cancer. Our results demonstrate that integrating multi-modal data helps achieve significant performance gains over image-only approaches. Future work will include incorporating radiologist imaging and time-series clinical data for further gains.

Funding

This research has no funding source.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to the privacy-preserving nature of the data, but are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] El Nahhas, O.S., van Treeck, M., Wölflein, G., Unger, M., Ligero, M., Lenz, T., Wagner, S.J., Hewitt, K.J., Khader, F., Foersch, S., et al.: From whole-slide image to biomarker prediction: end-to-end weakly supervised deep learning in computational pathology. *Nature Protocols* 20(1), 293–316 (2025)

[2] Hiwase, A.D., Ovenden, C.D., Kaukas, L.M., Finnis, M., Zhang, Z., O’Connor, S., Foo, N., Reddi, B., Wells, A.J., Ellis, D.Y.: Can rotational thermoelectrometry rapidly identify theragnostic targets in isolated traumatic brain injury? *Emergency Medicine Australasia* 37(1), e14480 (2025)

- [3] Paik, S., et al. (2024) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *New England Journal of Medicine*.
- [4] Wulczyn, E., et al. (2024) Deep learning-based survival prediction for multiple cancer types using histopathology images. *PLOS ONE*.
- [5] Courtiol, P., et al. (2019) Deep learning-based classification of mesothelioma improves prediction of patient outcome. *Nature Medicine*.
- [6] Bui, D.C., Vuong, T.T.L., Kwak, J.T. (2024). FALFormer: Feature-Aware Landmarks Self Attention for Whole-Slide Image Classification. In *MICCAI 2024*.
- [7] Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F., Jaume, G., Chen, B., Zhang, A., Shao, D., Song, A.H., Shaban, M., et al. (2024). Towards a General-Purpose Foundation Model for Computational Pathology. *Nature Medicine*.
- [8] Cui, M., Zhang, D.Y. (2021). Artificial Intelligence and Computational Pathology. *Laboratory Investigation*, 101(4), 412-422.
- [9] Luo, Y., Wang, S., Liu, J., Xiao, J., Xue, R., Zhang, Z., ... & Xie, Y. (2025). Pathohr: Breast cancer survival prediction on high-resolution pathological images. *arXiv preprint arXiv:2503.17970*.
- [10] Park, J., Witowski, J., Xu, Y., Trivedi, H., Gichoya, J., Brown-Mulry, B., ... & Geras, K. J. (2025). A Multi-Modal AI System for Screening Mammography: Integrating 2D and 3D Imaging to Improve Breast Cancer Detection in a Prospective Clinical Study. *arXiv preprint arXiv:2504.05636*.
- [11] Osuala, R., Lang, D. M., Riess, A., Kaissis, G., Szafranowska, Z., Skorupko, G., ... & Lekadir, K. (2024, October). Enhancing the utility of privacy-preserving cancer classification using synthetic data. In *Deep Breast Workshop on AI and Imaging for Diagnostic and Treatment Challenges in Breast Care* (pp. 54-64). Cham: Springer Nature Switzerland.
- [12] Høibo, M., Pedersen, A., Dale, V. G., Berget, S. M., Ytterhus, B., Lindskog, C., ... & Valla, M. (2025). Immunohistochemistry guided segmentation of benign epithelial cells, in situ lesions, and invasive epithelial cells in breast cancer slides. *PLoS One*, 20(7), e0328033.

Disclaimer/Publisher's Note: The perspectives, opinions, and data shared in all publications are the sole responsibility of the individual authors and contributors, and do not necessarily reflect the views of Sciences Force or the editorial team. Sciences Force and the editorial team disclaim any liability for potential harm to individuals or property resulting from the ideas, methods, instructions, or products referenced in the content.