Development and validation of an artificial intelligence system for triple-negative breast cancer identification and prognosis prediction: a multicentre retrospective study

Xiu-Ming Zhang,[a,l] Hua-Jun Zhou,[b,l] Qing Chen,[a,l] Xi Wang,[b,c,l] Yu-Juan Fu,[d,l] Cheng Jin,[b] Feng-Tao Zhou,[b] Jing-Ping Wang,[a] Qiu-Yu Cai,[a] Ji-Li Wang,[a] Bo Luo,[e] Mao-Tong Hu,[f] Cai-Yun Yao,[f] Xia Yang,[d] Ya-Li Xu,[g,***] Jing Zhang,[a,**] and Hao Chen[b,h,i,j,k,*]

[a]Department of Pathology, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, 310003, China
[b]Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong, China
[c]Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China
[d]Department of Pathology, Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou, 310016, China
[e]Department of Pathology, The Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430014, China
[f]Department of Pathology, Yiwu Central Hospital, Jinhua, 322000, China
[g]Department of Pathology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, 250021, China
[h]Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Hong Kong, China
[i]Division of Life Science, The Hong Kong University of Science and Technology, Hong Kong, China
[j]HKUST Shenzhen-Hong Kong Collaborative Innovation Research Institute, Futian, Shenzhen, 518048, China
[k]State Key Laboratory of Nervous System Disorders, The Hong Kong University of Science and Technology, Hong Kong, China

eClinicalMedicine
2025;89: 103557
Published Online xxx
https://doi.org/10.1016/j.eclinm.2025.103557

Summary

Background Triple-negative breast cancer (TNBC), recognised as the most aggressive subtype of breast cancer, has a high recurrence rate and poor treatment outcomes. Current diagnosis relies heavily on immunohistochemistry, which can be time-consuming and costly, while prognostic stratification remains limited by traditional clinicopathological features. This study aimed to develop and validate an artificial intelligence (AI)-powered TNBC identification and prognosis prediction (TRIP) system using haematoxylin and eosin (H&E)-stained pathology images.

Methods In this multicentre retrospective study, we analysed haematoxylin and eosin (H&E)-stained whole slide images (WSIs) of 2045 patients with breast cancer from The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China (FAH) between June 1, 2007 and December 31, 2022. Among these, 451 patients with TNBC had follow-up outcomes on disease-free survival and overall survival. Patients were excluded if they had other synchronous malignant neoplasms within five years or had previously received neoadjuvant chemotherapy. We developed and validated a deep learning system, TRIP, to classify TNBC versus other breast cancer subtypes and predict the cancer disease-free survival and overall survival on the FAH cohort. We employed the area under the receiver operating characteristic curve (AUC) to evaluate TNBC identification performance, and the concordance index (C-index) to evaluate the performance of disease-free survival and overall survival predictions. Beyond internal validation, the system was externally evaluated on independent retrospective cohorts from four tertiary hospitals in Eastern and Central China (Shandong Provincial Hospital [SDPH], Sir Run Run Shaw Hospital of Zhejiang University [SRRS], Yiwu Central Hospital [YWCH], and The Central Hospital of Wuhan [WHCH]) between June 26, 2013 and December 31, 2024, and The Cancer Genome Atlas (TCGA) dataset between January 1, 1988 and December 31, 2013, comprising 2793 cases for TNBC identification and 463 cases for prognosis prediction. Model interpretability was enhanced using pathology heatmaps, and multi-omics analysis was conducted to explore TNBC heterogeneity.

Findings The deep-learning model achieved an AUC of 0.980, 95% Confidence Interval (CI): 0.958–0.996, for identifying TNBC in the internal cohort, and AUCs of 0.916 (95% CI: 0.848–0.959), 0.936 (95% CI: 0.907–0.962), 0.860 (95% CI: 0.779–0.930), and 0.890 (95% CI: 0.841–0.929) in the SDPH, SRRS, WHCH, and TCGA external

*Corresponding author. Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong, China.
**Corresponding author. Department of Pathology, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, 310003, China.
***Corresponding author. Department of Pathology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, 250021, China.
E-mail addresses: jhc@cse.ust.hk (H. Chen), jzhang1989@zju.edu.cn (J. Zhang), xuyl@sdu.edu.cn (Y.-L. Xu).
[l]Contributed equally.
Translation For the Chinese translation of the abstract, see Supplementary Materials section.

validation cohorts, respectively. Moreover, it effectively predicted disease-free survival with a C-index of 0.747 ± 0.070 (95% CI: 0.617–0.852) in the internal cohort and C-indices of 0.731 ± 0.047 (95% CI: 0.623–0.839) and 0.732 ± 0.043 (95% CI: 0.621–0.840) in the SDPH, SRRS, YWCH cohorts combined (EXSURV), and the TCGA cohort, respectively, and overall survival with a C-index of 0.744 ± 0.075 (95% CI: 0.602–0.865) in the internal cohort and C-indices of 0.720 ± 0.034 (95% CI: 0.566–0.865) and 0.721 ± 0.030 (95% CI: 0.625–0.818) in the EXSURV and TCGA cohorts, respectively. Heatmaps revealed key histologic features of TNBC, including nuclear atypia, necrosis, and immune-cold microenvironments in aggressive cases, while lymphoplasmacytic infiltration indicated better prognosis. Multi-omics analysis identified three molecular subtypes with distinct immune and pro-tumour signalling profiles, supporting the TRIP system's prognostic accuracy in alignment with molecular evidence.

Interpretation TRIP is an effective and interpretable artificial intelligence system that demonstrates strong performance in identifying TNBC and predicting its disease-free survival and overall survival. However, the current findings are primarily derived from post-surgical tissue data and do not incorporate clinical variables, which limits the system's immediate applicability in pre-operative settings. To overcome these constraints, future prospective studies should be conducted to validate its clinical utility, particularly in pre-operative and broader real-world scenarios.

Funding National Natural Science Foundation of China; Hong Kong Innovation and Technology Commission; Zhejiang Provincial Natural Science Foundation of China; and National Health Commission Scientific Research Fund, China.

Keywords: Triple-negative breast cancer; Cancer diagnosis; Survival analysis; Artificial intelligence

Research in context

Evidence before this study

We searched PubMed with the search term ("triple-negative breast cancer") AND ("artificial intelligence" OR "deep learning") AND ("pathology" OR "histopathology") for published articles without language restrictions from database inception to May 7, 2025. We identified 177 studies, six of which applied artificial intelligence (AI) techniques for triple-negative breast cancer (TNBC) identification or prognosis prediction. However, the patient populations utilised to develop and validate AI models are typically less than two cohorts, comprising <600 TNBC patients altogether. Moreover, these studies mainly focused on building an AI model to address a specific task, instead of developing a system to facilitate the broader and deeper integration of AI into clinical practice.

Added value of this study

According to our literature search results, we developed TNBC identification and prognosis prediction (TRIP), an accurate, generalisable, and interpretable AI-powered system for TNBC identification and prognosis prediction. Addressing the gap in current AI research for TNBC, we rigorously validated TRIP across diverse TNBC populations from multiple centres in China as well as the TCGA dataset. To our knowledge, this study represents the largest dataset to date (>1000 TNBC patients) used for the development and validation of an AI system tailored to TNBC. Furthermore, it is the first AI system designed to simultaneously address multiple clinical tasks in TNBC, thereby facilitating its potential integration into clinical practice.

Implications of all the available evidence

This study built a publicly available AI system for the diagnosis and prognosis prediction of TNBC. Our system shows good performance in identifying patients with TNBC using haematoxylin and eosin (H&E)-stained pathology images, with the potential to streamline immunohistochemistry workflows and reduce pathologists' workload. In addition, its prognostic capability provides clinically relevant insights into disease progression, supporting more informed and timely clinical decision-making. Nevertheless, prospective trials are warranted to further validate its utility in real-world clinical practice and confirm its contribution to improving patient care.

Introduction

Triple-negative breast cancer (TNBC) accounts for about 15% of patients with breast cancer and is characterised by its aggressive nature, rapid growth, and high likelihood of metastasis and recurrence after treatment.[1] The 5-year survival rate of TNBC is about 75%, which is significantly lower than other breast cancer subtypes (>90%).[2] TNBC is defined as the tumours that lack the expression of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2), all of which are molecular targets of therapeutic agents.[3–5] In clinical

practice, the diagnosis of TNBC primarily relies on the assessment of ER and PR protein expression levels by immunohistochemistry (IHC), and of HER2 by IHC and/or fluorescence in situ hybridization (FISH).[6] However, the adoption of these molecular signature-based tests is limited by their high cost, long turn-around time, and requirement for suitable tissue samples. Thus, an alternative method that is practical, cost-effective, and utilises readily available samples for TNBC identification beyond receptor subtyping could be of great utility. Moreover, due to the high heterogeneity of TNBC, patients with the same tumour-lymph node-metastasis (TNM) stage may experience notably distinct clinical prognoses. Therefore, there is an urgent need to establish a model for predicting the prognosis of TNBC. Several prognostic biomarkers, such as stromal tumour-infiltrating lymphocytes (sTIL) and tumour-associated stroma (TAS), have been identified as influential for the survival of patients with TNBC.[7–9] However, given that numerous factors within the tumour microenvironment contribute to TNBC progression, solely depending on a limited number of biomarkers may not yield precise prognosis prediction.

Over the past decade, with tremendous advances in computing power, deep learning has become an effective tool in computational pathology, particularly for giga-pixel whole slide images (WSIs).[10–16] Significant progress has been made across various malignancies, such as lung, ovarian, liver, and colorectal cancers, with applications spanning diagnosis (e.g., tumour detection, subtyping, and grading) and prognosis (e.g., biomarker discovery, treatment response prediction, and survival analysis).[17–21] However, computational pathology for TNBC remains relatively underexplored. Existing studies are limited by small cohort sizes, overreliance on predefined biomarkers (e.g., stromal tumour-infiltrating lymphocytes), fragmented multi-stage pipelines, and lack of external validation, collectively hindering clinical translation.[22–27]

Previous efforts have primarily focused on molecular subtype prediction or semi-quantitative biomarker assessment. For instance, Gamble et al. proposed a two-stage pipeline combining patch-level convolutional neural networks with slide-level logistic regression, achieving moderate area under the receiver operating characteristic curves (AUCs) for ER (0.83), PR (0.72), and HER2 (0.58).[28] Jang et al. applied a weakly supervised framework (CLAM) to classify breast cancer subtypes but attained limited TNBC sensitivity (67.14%, n = 228).[26] In the prognostic domain, several studies explored tumour-infiltrating lymphocyte (TIL) features[22–25] or tumour-stroma ratios,[27] but their reliance on labour-intensive annotations or handcrafted features often led to moderate predictive performance and poor cross-cohort generalisability.

These limitations underscore the need for a robust, fully automated, and generalisable system capable of jointly identifying TNBC and predicting patient prognosis from standard haematoxylin and eosin (H&E)-stained WSIs. To address this gap, we developed and comprehensively validated a unified, end-to-end deep learning system, TRIP, that integrates a pathology foundation model with effective long-sequence modelling to extract clinically significant histological patterns across large-scale breast cancer cohorts from five top-tier hospitals and public datasets, as shown in Fig. 1. The TRIP system aims to support comprehensive TNBC management in routine pathology workflows, offering both diagnostic and prognostic utility in a single pipeline.

Methods

Study design and population

We defined the following inclusion and exclusion criteria to ensure cohort consistency across centres. Patients were eligible for inclusion if they had (1) histologically confirmed breast cancer, (2) available H&E-stained whole slide images, and (3) complete clinicopathological information. Patients were excluded if they had (1) other synchronous malignant neoplasms within five years or (2) had previously received neoadjuvant chemotherapy. All patients underwent IHC testing for ER, PR, and HER2 status, and those negative for all three markers were classified as triple-negative breast cancer. For specific analyses, survival cohorts were required to have complete follow-up data, and genetic analyses were performed only on patients with sufficient, qualified tissue specimens. A flow diagram of our inclusion/exclusion criteria is illustrated in Fig. S1.

In this multicentre retrospective study, we reviewed the data of 4898 patients with breast cancer from five medical centres and a public repository, and all patients underwent surgery. Among these, 2045 patients with breast cancer were treated at the First Affiliated Hospital of Zhejiang University (FAH) from June 1, 2007 to December 31, 2022, where 451 patients with TNBC had follow-up data on disease-free survival and overall survival; 421 patients with breast cancer were treated at Shandong Provincial Hospital (SDPH) from February 25, 2014 to December 31, 2023, where 79 patients with TNBC had follow-up data; 788 patients with breast cancer were treated at Sir Run Run Shaw Hospital of Zhejiang University (SRRS) from January 1, 2015 to December 31, 2024, where 140 patients with TNBC with follow-up data; 60 TNBC patients were treated at Yiwu Central Hospital (YWCH) from June 26, 2013 to August 1, 2022 with follow-up data; 522 patients with breast cancer were treated at The Central Hospital of Wuhan (WHCH) from January 1, 2020 to December 31, 2024 without follow-up data. The follow-up data from collaborating hospitals range from 28 to 97 months. Moreover, we also included 1062 patients with

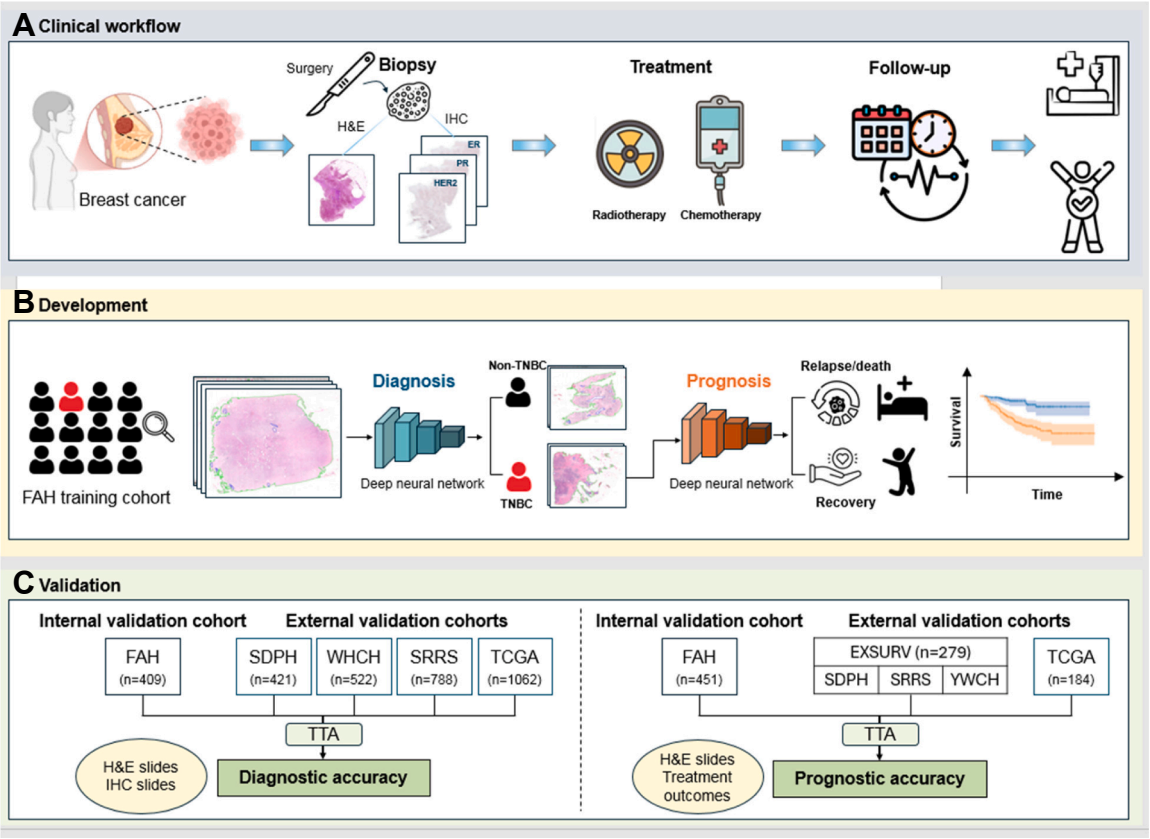

***Fig. 1:* Overview of study design**. A. Clinical workflow for the diagnosis, treatment, and prognosis for patients with triple-negative breast cancer (TNBC). Biopsy was performed during surgery for each patient. H&E– and IHC-stained slides are digitalised to whole slide images. B–C. Development and independent validation of an end-to-end deep learning system to identify TNBC patients and predict the disease-free survival and overall survival. The primary validation dataset, FAH, was collected at First Affiliated Hospital of Zhejiang University (Hangzhou, China). Four in-house external validation cohorts, SDPH, WHCH, SRRS, YWCH include data collected at Shandong Provincial Hospital (Jinan, China), Sir Run Run Shaw Hospital of Zhejiang University (Hangzhou, China), Yiwu Central Hospital (Yiwu, China), and The Central Hospital of Wuhan (Wuhan, China). TCGA dataset is from https://portal.gdc.cancer.gov/. EXSURV is a combined cohort of SDPH, SRRS, and YWCH. AUROC, area under the receiver operator characteristic curve. H&E, haematoxylin and eosin. IHC, immunohistochemistry. TNBC, triple-negative breast cancer. TTA, test-time adaptation.

breast cancer from The Cancer Genome Atlas (TCGA) dataset, diagnosed between January 1, 1988 and December 31, 2013, where 167 and 184 patients with TNBC had follow-up data on disease-free survival and overall survival, respectively.[29] A statistical summary of these cohorts is provided in Table S1.

We organised the above datasets into multiple cohorts for developing and validating the TRIP system. For TNBC identification, we randomly select 80% of patients (n = 1636) in the FAH cohort as the training cohort, while the rest of 20% (n = 409) patients constitute the internal validation cohort. Besides, the other four independent cohorts, i.e., SDPH, SRRS, WHCH, and TCGA cohorts, are employed as independent sets for external validation, separately. For prognosis prediction, we conducted five-fold cross-validation on the WSIs of 451 patients with TNBC from the FAH cohort. Given the limited number of TNBC patients with complete follow-up data in the SDPH, SRRS, and YWCH cohorts, we constructed a combined external validation cohort named EXSURV by aggregating patients from these three sources. This strategy increases the total sample size and statistical power for survival analysis, thereby mitigating the unreliability introduced by small individual cohorts. The TCGA cohort was also employed as an additional external validation set. We followed the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis Artificial Intelligence update (TRIPOD-AI) checklist to ensure best-practice reporting.

Outcomes

The primary objective was to use deep learning to identify TNBC among breast cancer populations from WSIs. TNBC is defined as IHC-negative for ER, PR, and

HER-2 (IHC score 0, 1+ or 2+ with negative HER-2 FISH confirmation). The secondary objective was to predict treatment outcomes for patients with TNBC, i.e., overall survival and disease-free survival. Overall survival is defined as the time from primary surgery to death from any cause (censored at the last follow-up). Disease-free survival was defined as the time from completion of primary treatment to recurrence, new primary cancer, or death (censoring at death or date of the last visit).

Deep learning system development

The proposed TNBC identification and prognosis prediction (TRIP) system is composed of two major modules, one responsible for TNBC discrimination from breast cancer populations, containing a TNBC classification model, and the other accountable for prognosis prediction, including a disease-free survival analysis model and an overall survival analysis model. Unlike previous approaches that design specialised architectures for individual tasks, we developed a unified multiple instance learning (MIL) framework shared across all tasks. This framework integrates a state-of-the-art pathology foundation model for patch-level feature extraction and a Mamba-based long-sequence modelling architecture for whole slide image analysis, as shown in Fig. S2. Following previous studies, we employed the CLAM toolkit[30] to divide each gigapixel WSI into non-overlapping 512 × 512 tissue patches, which were subsequently processed through feature extraction and aggregation to obtain slide-level representations. Motivated by the recent success of histopathology foundation models, we adopted GPFM,[31] the latest and most advanced model to date, as our feature extractor. Specifically, GPFM transforms the squared patches into a long sequence of patch embeddings without requiring additional stain normalisation techniques. Traditional MIL methods typically treat WSIs as unordered collections of patches and rely on permutation-invariant operations, such as selecting the most predictive instance (e.g., MaxMIL) or applying attention-based pooling (e.g., AttMIL[15]) to aggregate patch-level features into a slide-level representation. While these approaches can achieve reasonable performance, they do not explicitly model spatial dependencies or contextual relationships among distant patches. In contrast, our TRIP system introduces a bidirectional Mamba encoder that models patch sequences in both forward and reverse directions. Although the standard Mamba architecture enables efficient long-sequence modelling, its unidirectional nature may result in early sequence information being overlooked. By incorporating bidirectional processing, TRIP captures long-range dependencies across the entire slide and mitigates information loss. The bidirectional outputs are then fused and passed through a lightweight attention-based aggregator to generate task-specific predictions. To further improve robustness to domain shifts across validation cohorts, such as variations in staining, scanning, and site-specific biases, we adopted test-time adaptation (TTA)[32] to improve the model's generalisation ability by fine-tuning the parameters of normalisation layers while freezing all other parameters in the TRIP system. The binary cross-entropy loss and negative log-likelihood (NLL) loss with censorship were employed as the objective functions for TNBC identification and survival analysis, respectively. We trained our models for 20 epochs using the Adam optimiser with a learning rate of 1e-4. Of note, the final model was selected based on its best performance on the internal validation set across 20 training epochs. All experiments were conducted using the PyTorch framework, and further implementation details are described in the Supplementary Material.

Interpretability analysis

To assist with explainability, we generated saliency maps for the WSIs to investigate what areas were important for prediction. We extracted attention scores of all patches from the attention aggregator module in the TRIP model and applied the Min-Max normalisation to rescale them into the range of [0,1]. Afterwards, we generate the heatmaps by geographically stitching the normalised attention scores into a 2D map with a similar aspect ratio to the original WSI, where the points that correspond to background (i.e., non-tissue regions) are assigned zeros. These heatmaps are then overlapped with the thumbnails of the WSIs for visualisation. Based on these saliency maps, two board-certified pathologists (with 10 and 17 years of experience, respectively) manually examined the most discriminative patches to discover crucial patterns associated with TNBC identification and prognosis. Furthermore, to further demonstrate our findings, we extracted fourteen geometric and texture features between two compared samples using the analysis tool proposed by Zhao et al.[33]

Integrated multi-omics analysis

To confirm that our TRIP system's prognostic accuracy is in alignment with molecular evidence, 211 patients from the FAH cohort have undergone whole-genome RNA sequencing with details in our Supplementary Material. We performed integrated multi-omics analysis to characterise molecular heterogeneity in triple-negative breast cancer. First, we applied differential gene expression (DGE) analysis on the transcriptomic data. The high-risk and low-risk TNBC samples, stratified by the TRIP prognostic system, were compared using DESeq2 (v1.30.1) with thresholds of $|\log_2 \text{fold-change}| > 1$ and false discovery rate (FDR) < 0.05. Then, significant genes were visualised via a volcano plot, highlighting upregulated (red) and downregulated (blue) genes. Then, we applied gene set variation

analysis (GSVA) to quantify the enrichment scores of Hallmark gene sets (MSigDB v7.4) in each sample, followed by consensus clustering to define several molecular subgroups.

Statistical analysis

For TNBC identification, the performance of the model was assessed using AUC, sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), and F1 score, along with their 95% confidence intervals (CIs). Statistical comparisons of AUCs were performed using DeLong's test, while sensitivity and specificity were compared using the two-sided McNemar test to evaluate significant differences between models. As for survival analysis, the prognostic performance was quantified using Harrell's concordance index (C-index). Survival distributions between predicted high- and low-risk groups were compared using Kaplan–Meier analysis with log-rank testing.

Ethical approval

This study was approved by the Clinical Research Ethics Committee of the First Affiliated Hospital of Zhejiang University (Reference: 2024–1348). Four other hospitals, including Shandong Provincial Hospital, Sir Run Run Shaw Hospital of Zhejiang University, Yiwu Central Hospital, and The Central Hospital of Wuhan, have accepted the decision of the Clinical Research Ethics Committee of the First Affiliated Hospital of Zhejiang University. Informed consent was waived for the retrospective study. All public datasets were fully de-identified and publicly available, eliminating the need for additional ethical approvals.

Role of the funding source

The funders of this study had no roles in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. XMZ and HC had final responsibility for the decision to submit for publication.

Results

The clinicopathological characteristics of the patients in the primary dataset and the external validation datasets for TNBC identification and prognosis prediction are reported in Tables 1 and 2, respectively. For TNBC identification, the training cohort from FAH includes 1636 patients (mean age of 53.75), among which TNBC cases account for about 28.9% (n = 473), and the internal validation cohort contains 409 patients with 29.4% being TNBC. The external validation cohorts, i.e., SDPH, SRRS, WHCH, and TCGA, contain 421, 788, 522, and 1062 patients with mean ages of 51.83, 53.75, 57.08, and 58.42, respectively, and the corresponding patients with TNBC occupy 24.0%, 37.2%, 15.7%, and 17.4%, accordingly. In terms of laterality, the proportion of the left side and the right side is quite similar, and the occurrence rate of bilateral cases is quite marginal (i.e., less than 2%) across all cohorts. For TNBC prognosis prediction, the mean age of patients with TNBC in FAH, TCGA and EXSURV cohorts is 53.53, 53.62, and 55.39, respectively. The recurrence rates of FAH, TCGA and EXSURV are 15.7%, 7.9% and 11.4% while the mortality rates are 13.1%, 5.4% and 13.6%, accordingly. In the internal cohort, the median (IQR) follow-up time for DFS and OS was 71.8 (45.8, 96.3) and 74.2 (49.5, 97.1) months, respectively. In the combined external validation cohort, i.e., EXSURV, the median (IQR) follow-up time for DFS and OS was 50.0 (31.0, 68.5) and 51.0 (33.0, 69.0) months. In the public TCGA cohort, the median (IQR) follow-up time for DFS and OS was 28.2 (14.3, 59.4) and 29.8 (14.2, 62.0) months, respectively.

For the TNBC identification experiments, the results in five cohorts, i.e., the internal validation cohort FAH and four external cohorts: SDPH, SRRS, WHCH, and TCGA, are reported in Table 3. Our TRIP system with TTA showed high prediction efficacy in identifying TNBC among patients with breast cancer in the internal validation cohort (i.e., FAH), obtaining an AUC of 0.980 (95% CI 0.958–0.996), a sensitivity of 0.963 (95% CI 0.926–0.993), a specificity of 0.857 (95% CI 0.750–0.943), an accuracy of 0.934 (0.890–0.971), a PPV of 0.947 (0.897–0.980), a NPV of 0.897 (0.802–0.983) and a F1 score of 0.955 (0.922–0.980). When deployed to the external cohorts, the TRIP system with TTA demonstrated a good generalisability in various clinical scenarios, consistently achieving high predictive performance, e.g., AUCs of 0.916 (95% CI 0.848–0.959), 0.936 (95% CI 0.907–0.962), 0.860 (95% CI 0.779–0.930), and 0.890 (95% CI: 0.841–0.929) in the SDPH, SRRS, WHCH, and TCGA cohorts, respectively (Table 3, Fig. 2). In comparison to two state-of-the-art AI models, i.e., MaxMIL and AttMIL, TRIP showed prominent superiority, improving AUCs by 5.2% and 3.5% in absolute terms in the internal validation cohort (P-values <0.0001). Consistently, in external validation, TRIP boosted the AUCs by 5.0%–17.0% (mean 10.8%) compared to MaxMIL (P-values $\leq$0.0011) and 2.1%–7.5% (mean 4.7%) compared to AttMIL (P-values $\leq$0.083). To ensure a fair comparison, we compared the specificities of the three deep learning models by maintaining similar sensitivities. Within the internal validation cohort, when sensitivity was kept at 96.3%, TRIP achieved a much higher specificity (85.7% [95% CI: 75.0%–94.3%]) than MaxMIL (67.0%, P-values <0.0001) and AttMIL (70.5%, P-values = 0.0001), with 15.2%–18.7% improvement. In the external validation cohorts, while maintaining a similar sensitivity to other models, TRIP showed a specificity of 76.8%–85.1%, which significantly outperformed the MaxMIL model

Variables	Internal cohort						External cohorts											
	FAH training cohort (n = 1636)			FAH validation cohort (n = 409)			SDPH (n = 421)			SRRS (n = 788)			WHCH (n = 522)			TCGA (n = 1062)		
	Non-TNBC	TNBC	P-value	Non-TNBC	TNBC	P-value	Non-TNBC	TNBC	P-value	Non-TNBC	TNBC	P-value	Non-TNBC	TNBC	P-value	Non-TNBC	TNBC	P-value
No. (%)	1163 (71.1%)	473 (28.9%)		293 (71.6%)	116 (28.4%)		320 (76.0%)	101 (24.0%)		495 (62.8%)	293 (37.2%)		440 (84.3%)	82 (15.7%)		877 (82.6%)	185 (17.4%)	
Age			0.012			0.24			0.15			0.048			0.031			0.0009
Mean (SD)	53.16 (11.50)	54.53 (11.12)		52.99 (11.99)	54.76 (12.70)		51.69 (10.53)	53.01 (9.14)		53.21 (11.22)	54.65 (11.39)		57.52 (12.22)	54.70 (13.45)		59.07 (13.33)	55.39 (12.22)	
Median (Min, Max)	52 (23, 92)	54 (24, 93)		53 (22. 86)	55 (29, 85)		51 (28, 82)	53 (32, 75)		52 (25, 94)	54.5 (22, 92)		57 (27, 91)	52.5 (30, 90)		59 (26, 90)	53 (29, 90)	
Family history of breast cancer			0.017			0.0007			0.90			0.61			0.62			–
No	1062 (91.3%)	448 (94.7%)		289 (98.6%)	101 (87.1%)		291 (90.9%)	91 (90.1%)		451 (91.1%)	272 (92.8%)		399 (90.7%)	74 (90.2%)		0 (0%)	0 (0%)	
Yes	79 (6.8%)	17 (3.6%)		1 (0.3%)	13 (11.2%)		19 (5.9%)	7 (6.9%)		29 (5.9%)	14 (4.8%)		26 (5.9%)	3 (3.7%)		0 (0%)	0 (0%)	
Unknown	22 (1.9%)	8 (1.7%)		3 (1.0%)	2 (1.7%)		10 (3.1%)	3 (3.0%)		15 (3.0%)	7 (2.4%)		15 (3.4%)	5 (6.1%)		877 (100%)	185 (100%)	
Histological type			0.12			0.72			0.11			0.022			0.46			–
Not otherwise specified	1094 (94.1%)	456 (96.4%)		279 (95.2%)	111 (95.7%)		295 (92.2%)	99 (98.0%)		482 (97.4%)	275 (93.9%)		379 (86.1%)	69 (84.1%)		0 (0%)	0 (0%)	
Infiltrating lobular carcinoma	14 (1.2%)	2 (0.4%)		1 (0.3%)	1 (0.9%)		2 (0.6%)	0 (0%)		0 (0%)	2 (0.7%)		16 (3.6%)	1 (1.2%)		0 (0%)	0 (0%)	
Others	55 (4.7%)	15 (3.2%)		13 (4.4%)	4 (3.4%)		23 (7.2%)	2 (2.0%)		13 (2.6%)	16 (5.5%)		45 (10.2%)	12 (14.6%)		877 (100%)	185 (100%)	
Laterality			0.20			0.28			0.68			0.037			0.60			–
Bilateral	3 (0.3%)	4 (0.8%)		0 (0%)	1 (0.9%)		0 (0%)	0 (0%)		8 (1.6%)	0 (0%)		3 (0.7%)	0 (0%)		0 (0%)	0 (0%)	
Left	606 (52.1%)	236 (49.9%)		152 (51.9%)	61 (52.6%)		190 (59.4%)	57 (56.4%)		239 (48.3%)	151 (51.5%)		236 (53.6%)	41 (50.0%)		0 (0%)	0 (0%)	
Right	554 (47.6%)	233 (49.3%)		141 (48.1%)	54 (46.5%)		130 (40.6%)	44 (43.6%)		248 (50.1%)	126 (43.0%)		200 (45.5%)	41 (50.0%)		0 (0%)	0 (0%)	
Unknown	0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	16 (5.5%)		1 (0.2%)	0 (0%)		877 (100%)	185 (100%)	
Grade			0.0005			0.0008			0.0004			0.0008			0.0006			–
I	64 (5.5%)	1 (0.2%)		17 (5.8%)	1 (0.9%)		17 (5.3%)	0 (0%)		28 (5.7%)	3 (1.0%)		28 (6.4%)	0 (0%)		0 (0%)	0 (0%)	
II	619 (53.2%)	98 (20.7%)		148 (50.5%)	28 (24.1%)		205 (64.1%)	32 (31.7%)		268 (54.1%)	53 (18.1%)		273 (62.0%)	16 (19.5%)		0 (0%)	0 (0%)	
III	480 (41.3%)	374 (79.1%)		128 (43.7%)	87 (75.0%)		98 (30.6%)	69 (68.3%)		187 (37.8%)	217 (74.1%)		132 (30.0%)	65 (79.3%)		0 (0%)	0 (0%)	
Unknown	0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)		12 (2.4%)	20 (6.8%)		7 (1.6%)	1 (1.2%)		877 (100%)	185 (100%)	
Stage			0.0011			0.97			0.048			0.15			0.93			0.0008
I	506 (43.5%)	186 (39.3%)		123 (42.0%)	47 (40.5%)		101 (31.6%)	34 (33.6%)		207 (41.8%)	120 (41.0%)		158 (35.9%)	28 (34.1%)		145 (16.5%)	32 (17.3%)	
II	498 (42.8%)	245 (51.8%)		133 (45.4%)	54 (46.6%)		136 (42.5%)	45 (44.6%)		216 (43.6%)	127 (43.3%)		273 (62.0%)	53 (64.6%)		479 (54.6%)	124 (67.0%)	
III	150 (12.9%)	40 (8.5%)		35 (11.9%)	14 (12.1%)		56 (17.5%)	12 (11.9%)		41 (8.3%)	28 (9.6%)		8 (1.8%)	1 (1.2%)		221 (25.2%)	23 (12.4%)	
IV	0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	0 (0%)		0 (0%)	3 (1.0%)		1 (0.2%)	0 (0%)		14 (1.6%)	5 (2.7%)	

(Table 1 continues on next page)

Variables	Internal cohort			External cohorts											
	FAH training cohort (n = 1636)			FAH validation cohort (n = 409)			SDPH (n = 421)			SRRS (n = 788)			WHCH (n = 522)			TCGA (n = 1062)		
	Non-TNBC	TNBC	P-value	Non-TNBC	TNBC	P-value	Non-TNBC	TNBC	P-value	Non-TNBC	TNBC	P-value	Non-TNBC	TNBC	P-value	Non-TNBC	TNBC	P-value
(Continued from previous page)																		
Unknown	9 (0.8%)	3 (0.6%)		2 (0.7%)	1 (0.9%)		27 (8.4%)	10 (9.9%)		31 (6.3%)	15 (5.1%)		0 (0%)	0 (0%)		18 (2.1%)	1 (0.6%)	
Lymph node metastasis			0.0040			0.079			0.012			0.0007			0.031			0.0004
Yes	406 (34.9%)	130 (27.5%)		99 (33.8%)	28 (24.1%)		134 (41.9%)	27 (26.7%)		163 (32.9%)	59 (20.1%)		70 (15.9%)	5 (6.1%)		474 (54.0%)	67 (36.2%)	
No	750 (64.5%)	341 (72.1%)		193 (65.9%)	87 (75.0%)		162 (50.6%)	64 (63.4%)		301 (60.8%)	219 (74.7%)		370 (84.1%)	77 (93.9%)		385 (43.9%)	117 (63.3%)	
Unknown	7 (0.6%)	2 (0.4%)		1 (0.3%)	1 (0.9%)		26 (8.1%)	10 (9.9%)		31 (6.3%)	15 (5.1%)		0 (0%)	0 (0%)		18 (2.1%)	1 (0.5%)	

FAH, the First Affiliated Hospital of Zhejiang University. SDPH, Shandong Provincial Hospital. SRRS, Sir Run Run Shaw Hospital of Zhejiang University. WHCH, The Central Hospital of Wuhan. TCGA, The Cancer Genome Atlas. No., number. SD, standard deviation. Min, minimum. Max, maximum.

Table 1: **Characteristics of patients in the training, internal, and external validation cohorts for triple-negative breast cancer identification.**

(43.6%–74.7%) and AttMIL model (64.3%–80.5%), with an average improvement of 24.6% and 11.8%, respectively. Of note, the TTA strategy enhanced the generalisation capability of the TRIP model, with a mean increase of 1.0% in AUC (P-values $\leq$0.35) and 2.0% in specificity in the external validation cohorts. We also incorporated TTA into both AttMIL and MaxMIL for TNBC identification in Table S2. Although TTA improved the performance of both models, their results consistently lagged behind those of our proposed TRIP system, highlighting the added effectiveness of TRIP beyond test-time adaptation alone.

Following the identification task, we assessed the prognostic capability of TRIP across the internal and external datasets in Table 4. In the FAH cohort, TRIP with TTA achieved remarkable C-index values of 0.747 (SD: 0.070) for disease-free survival (DFS) and 0.744 (SD: 0.075) for overall survival (OS). When applied to the external validation cohorts, a modest decline in performance was observed, ranging from 1.5% to 1.56% for DFS prediction and 2.4% to 2.3% for OS prediction. Specifically, TRIP yielded C-index values of 0.731 (SD: 0.047) on EXSURV and 0.732 (SD: 0.043) on TCGA for DFS prediction, and 0.720 (SD: 0.034) on EXSURV and 0.721 (SD: 0.030) on TCGA for OS prediction. In comparison, the two existing models exhibited inferior performance. On the internal FAH cohort, the MaxMIL model achieved C-index values of 0.589 (SD: 0.110) in DFS prediction and 0.601 (SD: 0.105) in OS prediction, while the AttMIL model performed slightly better, with values of 0.679 (SD: 0.052) and 0.698 (SD: 0.034), accordingly. However, both models demonstrated substantial generalisation gaps when evaluated on external datasets. For instance, the MaxMIL model failed to generate valid predictions on the TCGA cohort (C-index <0.5), and the AttMIL model showed marked performance degradation—recording declines of 4.7% (EXSURV) and 23.7% (TCGA) for DFS prediction, and 18.7% (EXSURV) and 31.9% (TCGA) for OS prediction. Notably, incorporating TTA significantly improved the generalisation of AI models, as shown in Table S3. With TTA, the average C-index values of TRIP increased from 0.682 to 0.737 for DFS prediction and from 0.688 to 0.728 for OS prediction. Importantly, the TRIP + TTA model achieved C-index values exceeding 0.720 across all external cohorts and outperformed other two models with TTA, highlighting its robustness and adaptability. In addition, Kaplan–Meier survival analyses in Fig. 3 showed that TRIP effectively stratified patients into high- and low-risk groups with statistically significant differences (log-rank P-values <0.0033) across all datasets. We also exhibited the Kaplan–Meier survival curves for both AttMIL and MaxMIL in Fig. S3. While both models demonstrated effective patient stratification in the internal FAH cohort, their performance notably declined

Variables	Internal cohort	External cohorts				
	FAH (n = 451)	SDPH (n = 79)	SRRS (n = 140)	YWCH (n = 60)	EXSURV (n = 279)	TCGA (n = 184)
Age						
Mean (SD)	53.53 (11.58)	51.54 (9.63)	54.54 (10.01)	54.23 (13.63)	53.62 (12.56)	55.39 (12.22)
Median (Min, Max)	53 (22, 93)	53 (32, 75)	54 (25, 79)	54.5 (25, 88)	54 (25, 88)	58 (26, 90)
Family history of breast cancer						
No	422 (93.6%)	73 (92.4%)	132 (94.3%)	54 (90.0%)	259 (92.8%)	0 (0%)
Yes	23 (5.1%)	4 (5.1%)	6 (4.3%)	4 (6.7%)	14 (5.0%)	0 (0%)
Unknown	6 (1.3%)	2 (2.5%)	2 (1.4%)	2 (3.3%)	6 (2.2%)	184 (100.0%)
Histological type						
Not Otherwise Specified	432 (95.8%)	77 (97.5%)	124 (88.6%)	50 (83.3%)	251 (90.0%)	0 (0%)
Infiltrating lobular carcinoma	3 (0.7%)	0 (0%)	2 (1.4%)	0 (0%)	2 (0.7%)	0 (0%)
Others	16 (3.5%)	2 (2.5%)	14 (10.0%)	10 (16.7%)	26 (9.3%)	184 (100.0%)
Laterality						
Bilateral	4 (0.9%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Left	232 (51.4%)	49 (62.0%)	75 (53.6%)	24 (40.0%)	148 (53.0%)	0 (0%)
Right	215 (47.7%)	30 (38.0%)	65 (46.4%)	36 (60.0%)	131 (47.0%)	0 (0%)
Unknown	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	184 (100.0%)
Grade						
I	1 (0.2%)	0 (0%)	2 (1.4%)	1 (1.7%)	3 (1.1%)	0 (0%)
II	95 (21.1%)	25 (31.6%)	26 (18.6%)	17 (28.3%)	68 (24.4%)	0 (0%)
III	355 (78.7%)	54 (68.4%)	112 (80.0%)	38 (63.3%)	204 (73.1%)	0 (0%)
Unknown	0 (0%)	0 (0%)	0 (0%)	4 (6.7%)	4 (1.4%)	184 (100.0%)
Stage						
I	189 (41.9%)	27 (34.2%)	66 (47.1%)	25 (41.7%)	118 (42.3%)	32 (17.4%)
II	223 (49.4%)	35 (44.3%)	56 (40.0%)	28 (46.7%)	119 (42.6%)	124 (67.4%)
III	36 (8.0%)	9 (11.4%)	15 (10.7%)	7 (11.7%)	31 (11.1%)	23 (12.5%)
IV	0 (0%)	0 (0%)	3 (2.2%)	0 (0%)	3 (1.1%)	4 (2.2%)
Unknown	3 (0.7%)	8 (10.1%)	0 (0%)	0 (0%)	8 (2.9%)	1 (0.5%)
Lymph node metastasis						
Yes	114 (25.3%)	18 (22.8%)	12 (8.6%)	16 (26.7%)	46 (16.5%)	67 (36.5%)
No	334 (74.0%)	53 (67.1%)	128 (91.4%)	44 (73.3%)	225 (80.6%)	116 (63.0%)
Unknown	3 (0.7%)	8 (10.1%)	0 (0%)	0 (0%)	8 (2.9%)	1 (0.5%)
Treatment						
Mastectomy	149 (33.0%)	6 (7.6%)	12 (8.6%)	11 (18.3%)	29 (10.4%)	0 (0%)
Breast-conserving surgery	123 (27.3%)	4 (5.1%)	60 (42.9%)	14 (23.3%)	78 (28.0%)	0 (0%)
Modified radical mastectomy	176 (39.0%)	69 (87.3%)	68 (48.5%)	34 (56.7%)	171 (61.3%)	0 (0%)
Nipple-sparing mastectomy	3 (0.7%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Unknown	0 (0%)	0 (0%)	0 (0%)	1 (1.7%)	1 (0.3%)	184 (100.0%)
# of recurrence follow-up	451 (100%)	79 (100%)	140 (100%)	60 (100%)	279 (100%)	167 (90.8%)
Disease free survival in months: Median (IQR)	71.8 (45.8, 96.3)	85.0 (40.0, 97.0)	41.5 (28.0, 58,0)	55.0 (40.0, 75.8)	50.0 (31.0, 68.5)	28.2 (14.3, 59.4)
Recurrence	71 (15.7%)	3 (3.8%)	12 (8.6%)	7 (11.7%)	22 (7.9%)	19 (11.4%)
# of survival follow-up	451 (100%)	79 (100%)	140 (100%)	60 (100%)	279 (100%)	184 (100%)
Overall survival in months: Median (IQR)	74.2 (49.5, 97.1)	85.0 (42.5, 97.0)	41.5 (28.0, 58.0)	52.5 (35.8, 72.3)	51.0 (33.0, 69.0)	29.8 (14.2, 62.0)
Death	59 (13.1%)	2 (2.5%)	8 (5.7%)	5 (8.3%)	15 (5.4%)	25 (13.6%)

FAH, the First Affiliated Hospital of Zhejiang University. SDPH, Shandong Provincial Hospital. SRRS, Sir Run Run Shaw Hospital of Zhejiang University. YWCH, Yiwu Central Hospital. TCGA, The Cancer Genome Atlas. EXSURV, a combined external validation cohort of SDPH, SRRS, and YWCH cohorts. SD, standard deviation. Min, minimum. Max, maximum. #, number. IQR, interquartile range.

Table 2: **Characteristics of patients in the training, internal, and external validation cohorts for triple-negative breast cancer prognosis prediction.**

in the external EXSURV and TCGA cohorts, indicating limited generalisability across independent datasets.

To further demonstrate the robustness of TRIP, we benchmarked GPFM against two representative alternatives, UNI[34] and PLIP,[35] with comparative results provided in Tables S4 and S5. While performance varies with the choice of backbone, TRIP maintains robust performance across all tested representations, demonstrating its generalisability across diverse settings. In addition, we also conducted experiments on sub-group analysis, including age (using a threshold of 50-year-olds), HER2 status (0 and low groups), and untreated subgroups in the FAH cohort, as shown in Table S6 and Fig. S4. Our TRIP system consistently showed good performance on these sub-groups analyses, demonstrating its generalisability.

Model	AUC	P-value	Sensitivity	P-value	Specificity	P-value	Accuracy	PPV	NPV	F1
Internal FAH cohort										
MaxMIL	0.928 (0.871–0.970)	<0.0001	0.963 (0.928–0.993)	1.000	0.670 (0.482–0.789)	<0.0001	0.883 (0.804–0.929)	0.885 (0.784–0.934)	0.872 (0.728–0.973)	0.923 (0.868–0.955)
AttMIL	0.945 (0.904–0.976)	<0.0001	0.963 (0.930–0.993)	1.000	0.705 (0.583–0.842)	0.0001	0.892 (0.834–0.941)	0.897 (0.829–0.943)	0.878 (0.753–0.978)	0.929 (0.883–0.963)
TRIP	0.974 (0.951–0.992)	0.018	0.963 (0.926–0.994)	1.000	0.839 (0.733–0.925)	0.32	0.929 (0.890–0.971)	0.941 (0.896–0.974)	0.895 (0.804–0.980)	0.952 (0.925–0.981)
TRIP + TTA	0.980 (0.958–0.996)	Reference	0.963 (0.926–0.993)	Reference	0.857 (0.750–0.943)	Reference	0.934 (0.890–0.971)	0.947 (0.897–0.980)	0.897 (0.802–0.983)	0.955 (0.922–0.980)
External SDPH cohort										
MaxMIL	0.746 (0.646–0.815)	<0.0001	0.844 (0.764–0.899)	1.000	0.436 (0.288–0.592)	<0.0001	0.746 (0.682–0.803)	0.826 (0.764–0.890)	0.468 (0.266–0.614)	0.835 (0.781–0.876)
AttMIL	0.850 (0.761–0.909)	<0.0001	0.844 (0.778–0.900)	1.000	0.653 (0.510–0.796)	<0.0001	0.798 (0.720–0.862)	0.885 (0.818–0.936)	0.569 (0.409–0.711)	0.864 (0.811–0.907)
TRIP	0.909 (0.847–0.950)	0.081	0.844 (0.781–0.917)	1.000	0.802 (0.673–0.926)	0.18	0.834 (0.774–0.898)	0.931 (0.881–0.971)	0.618 (0.482–0.750)	0.885 (0.838–0.932)
TRIP + TTA	0.916 (0.848–0.959)	Reference	0.844 (0.774–0.909)	Reference	0.851 (0.724–0.951)	Reference	0.843 (0.791–0.903)	0.931 (0.896–0.983)	0.618 (0.476–0.769)	0.891 (0.851–0.932)
External SRRS cohort										
MaxMIL	0.886 (0.834–0.926)	<0.0001	0.877 (0.829–0.917)	1.000	0.747 (0.664–0.815)	<0.0001	0.829 (0.783–0.872)	0.854 (0.799–0.905)	0.782 (0.699–0.853)	0.865 (0.823–0.900)
AttMIL	0.915 (0.873–0.949)	0.0004	0.877 (0.833–0.923)	1.000	0.805 (0.726–0.878)	0.061	0.850 (0.805–0.895)	0.884 (0.832–0.921)	0.795 (0.704–0.876)	0.880 (0.845–0.918)
TRIP	0.927 (0.894–0.956)	0.0027	0.877 (0.819–0.926)	1.000	0.829 (0.761–0.892)	0.17	0.859 (0.810–0.893)	0.897 (0.855–0.943)	0.799 (0.726–0.889)	0.887 (0.842–0.919)
TRIP + TTA	0.936 (0.907–0.962)	Reference	0.877 (0.827–0.920)	Reference	0.840 (0.771–0.913)	Reference	0.863 (0.827–0.909)	0.902 (0.855–0.944)	0.801 (0.721–0.867)	0.889 (0.855–0.930)
External WHCH cohort										
MaxMIL	0.766 (0.677–0.882)	0.0011	0.805 (0.750–0.873)	1.000	0.524 (0.354–0.758)	0.0014	0.761 (0.699–0.826)	0.901 (0.844–0.962)	0.333 (0.208–0.474)	0.850 (0.801–0.895)
AttMIL	0.834 (0.762–0.904)	0.083	0.805 (0.743–0.863)	1.000	0.683 (0.468–0.843)	0.31	0.785 (0.726–0.841)	0.932 (0.889–0.971)	0.394 (0.270–0.523)	0.863 (0.816–0.901)
TRIP	0.853 (0.773–0.922)	0.35	0.805 (0.751–0.866)	1.000	0.756 (0.581–0.892)	0.85	0.797 (0.751–0.856)	0.947 (0.905–0.981)	0.419 (0.270–0.557)	0.870 (0.835–0.909)
TRIP + TTA	0.860 (0.779–0.930)	Reference	0.805 (0.749–0.860)	Reference	0.768 (0.582–0.926)	Reference	0.799 (0.741–0.852)	0.949 (0.909–0.986)	0.423 (0.310–0.555)	0.871 (0.832–0.907)
External TCGA cohort										
MaxMIL	0.773 (0.713–0.847)	<0.0001	0.830 (0.788–0.866)	1.000	0.568 (0.443–0.696)	<0.0001	0.784 (0.739–0.828)	0.901 (0.861–0.933)	0.413 (0.327–0.492)	0.864 (0.830–0.892)
AttMIL	0.815 (0.760–0.863)	<0.0001	0.830 (0.788–0.873)	1.000	0.643 (0.520–0.742)	<0.0001	0.798 (0.755–0.832)	0.917 (0.886–0.947)	0.444 (0.345–0.539)	0.871 (0.841–0.896)
TRIP	0.872 (0.829–0.924)	0.0012	0.830 (0.790–0.879)	1.000	0.789 (0.678–0.884)	0.092	0.823 (0.790–0.863)	0.949 (0.924–0.972)	0.495 (0.406–0.646)	0.886 (0.861–0.910)
TRIP + TTA	0.890 (0.841–0.929)	Reference	0.830 (0.791–0.866)	Reference	0.798 (0.689–0.877)	Reference	0.824 (0.776–0.860)	0.947 (0.921–0.973)	0.497 (0.407–0.603)	0.887 (0.856–0.908)

Numbers in parentheses indicate 95% confidence intervals (95% CIs). The P-values are employed for a comparison with artificial intelligence (AI) models. We adjust the thresholds of AI models to report the same sensitivity score for a fair comparison. Statistical comparisons of area under the receiver operating characteristic curves (AUCs) were performed using DeLong's test, while sensitivity and specificity were compared using the two-sided McNemar test to evaluate significant differences between models. FAH, the First Affiliated Hospital of Zhejiang University. SDPH, Shandong Provincial Hospital. SRRS, Sir Run Run Shaw Hospital of Zhejiang University. WHCH, The Central Hospital of Wuhan. TCGA, The Cancer Genome Atlas. PPV, positive predictive value. NPV, negative predictive value.

Table 3: **Model performance on triple-negative breast cancer (TNBC) identification.**

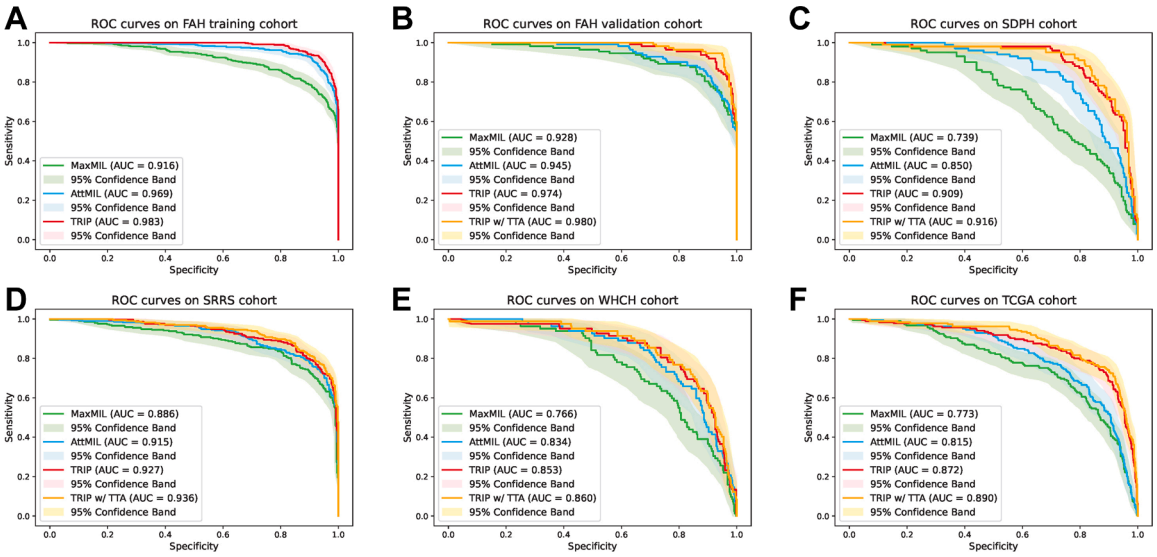

Fig. 2: **Receiver operator characteristic (ROC) curves for triple-negative breast cancer identification**. We compared our TRIP system with/without test-time adaptation (TTA) with two existing multiple instance learning models in the training cohort (A), the internal validation cohort (B), and four external validation cohorts (C–F).

We illustrated representative patches within WSIs based on our TRIP system in Fig. 4. For TNBC identification, we found that TNBC tumour cells exhibited marked nuclear atypia with pleomorphic, irregular nuclei containing vesicular chromatin and prominent nucleoli. Moreover, frequent mitotic figures were observed, and the neoplastic cells predominantly arranged in solid nests or sheets, lacking glandular/tubular structures. The stroma demonstrates prominent fibrotic proliferation with collagenous scar formation. Central tumour necrosis is also commonly present. For prognosis prediction, TNBC cases with poor prognosis compared to those with favourable prognosis typically demonstrate more prominent desmoplastic stroma, exhibit more aggressive infiltrative growth patterns characterised by interdigitating tumour margins, and show less lymphocytic and plasma cell infiltration in the interstitium. In addition, tumour necrosis is also significantly more common in the poor-prognosis group. The above findings are validated by the quantitative comparison, using the geometric and texture features developed previously by Yang et al.,[36] and in concordance with previous studies.[37,38]

Finally, to further validate the prognostic predictions, we conducted integrative multi-omics

Model	Disease-free survival				Overall survival			
	FAH	EXSURV	TCGA	Mean	FAH	EXSURV	TCGA	Mean
MaxMIL	0.589 ± 0.110	0.525 ± 0.035	0.490 ± 0.025	0.534	0.601 ± 0.105	0.485 ± 0.104	0.487 ± 0.025	0.524
95% CI	(0.425–0.743)	(0.394–0.649)	(0.341–0.601)	–	(0.406–0.771)	(0.326–0.627)	(0.374–0.600)	–
P-value	<0.0001	<0.0001	<0.0001	–	<0.0001	<0.0001	<0.0001	–
AttMIL	0.679 ± 0.052	0.632 ± 0.048	0.442 ± 0.017	0.584	0.698 ± 0.034	0.511 ± 0.055	0.379 ± 0.032	0.529
95% CI	(0.521–0.806)	(0.487–0.733)	(0.343–0.587)	–	(0.553–0.795)	(0.344–0.685)	(0.270–0.486)	–
P-value	<0.0001	<0.0001	<0.0001	–	<0.0001	<0.0001	<0.0001	–
TRIP	0.713 ± 0.051	0.673 ± 0.042	0.661 ± 0.021	0.682	0.718 ± 0.044	0.665 ± 0.049	0.682 ± 0.031	0.688
95% CI	(0.560–0.849)	(0.532–0.800)	(0.526–0.768)	–	(0.580–0.823)	(0.494–0.827)	(0.581–0.782)	–
P-value	<0.0001	<0.0001	<0.0001	–	<0.0001	<0.0001	<0.0001	–
TRIP + TTA	0.747 ± 0.070	0.731 ± 0.047	0.732 ± 0.043	0.737	0.744 ± 0.075	0.720 ± 0.034	0.721 ± 0.030	0.728
95% CI	(0.617–0.852)	(0.623–0.839)	(0.621–0.840)	–	(0.602–0.865)	(0.566–0.865)	(0.625–0.818)	–
P-value	Reference	Reference	Reference	–	Reference	Reference	Reference	–

FAH, the First Affiliated Hospital of Zhejiang University. EXSURV, a combined external validation cohort of SDPH, SRRS, and Yiwu Central Hospital cohorts. TCGA, The Cancer Genome Atlas. CI, confidence interval.

Table 4: **Model performance on disease-free and overall survival analysis for triple-negative breast cancer (TNBC).**

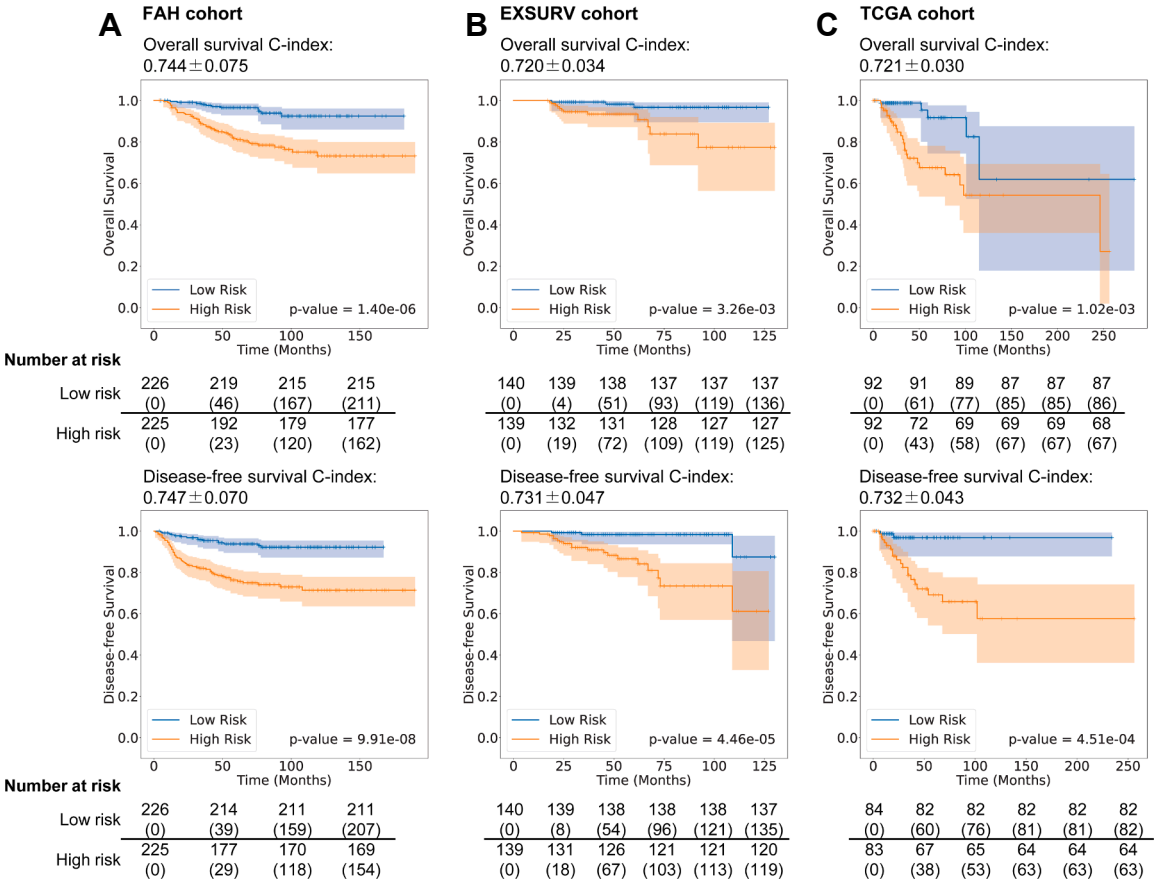

***Fig. 3*: Kaplan–Meier analysis for overall survival analysis (the upper panel) and disease-free survival analysis (the lower panel).** We illustrate the results of our TNBC identification and prognosis prediction (TRIP) system with test-time adaptation (TTA) on the internal validation cohort (A), and the external validation cohorts (B–C). The number of censored individuals in each group are in brackets.

analyses. Integration of multi-omics profiling identified three distinct clusters (i.e., C1–C3) each with unique transcriptional and pathway activation profiles. Between high- and low-risk groups stratified by the prognosis prediction from TRIP, 84 upregulated and 32 downregulated genes were identified (Fig. 5A) via differential gene expression analysis. Then, GSVA-based pathway activity (Fig. 5B) revealed that C2 exhibited high activity in pathways like *HALLMARK_ALLOGRAFT_REJECTION*, *HALLMARK_INTERFERON_GAMMA_RESPONSE*, and *HALLMARK_INTERFERON_ALPHA_RESPONSE*, which are pivotal in enhancing the immune system, consistent with the immunomodulatory subtype of TNBC and correlating with better prognosis. Moreover, C3 exhibits TNBC mesenchymal subtype-associated epithelial–mesenchymal transition (EMT) features and is correlated with poor prognosis, while the *HALLMARK_ANDROGEN_RESPONSE* pathway in C1 potentially linked to the luminal androgen receptor (LAR) subtype. Furthermore, as demonstrated by the Kaplan–Meier analysis (Fig. 5C), cluster C2 is associated with a significantly better prognosis compared to clusters C1 and C3 (P-values <0.05).[39] These findings confirm that the prognostic predictions of our TRIP system are congruent with existing molecular findings in the literature.[40,41]

Discussion

In this multicentre retrospective study, we developed and validated TRIP, an end-to-end deep learning system for the automated identification and prognosis prediction of TNBC using routine H&E-stained WSIs. Leveraging over 4000 breast cancer cases, including more than 1000 patients with TNBC, from five tertiary hospitals of China and the TCGA dataset, TRIP demonstrates high diagnostic accuracy (AUC = 0.980 on internal, 0.853–0.927 on external cohorts) and consistent prognostic stratification (C-index = 0.720–0.747). These results underscore TRIP's robustness and generalisability across diverse populations.

Compared to previous WSI-based studies that relied on patch-level classifiers, handcrafted features, or

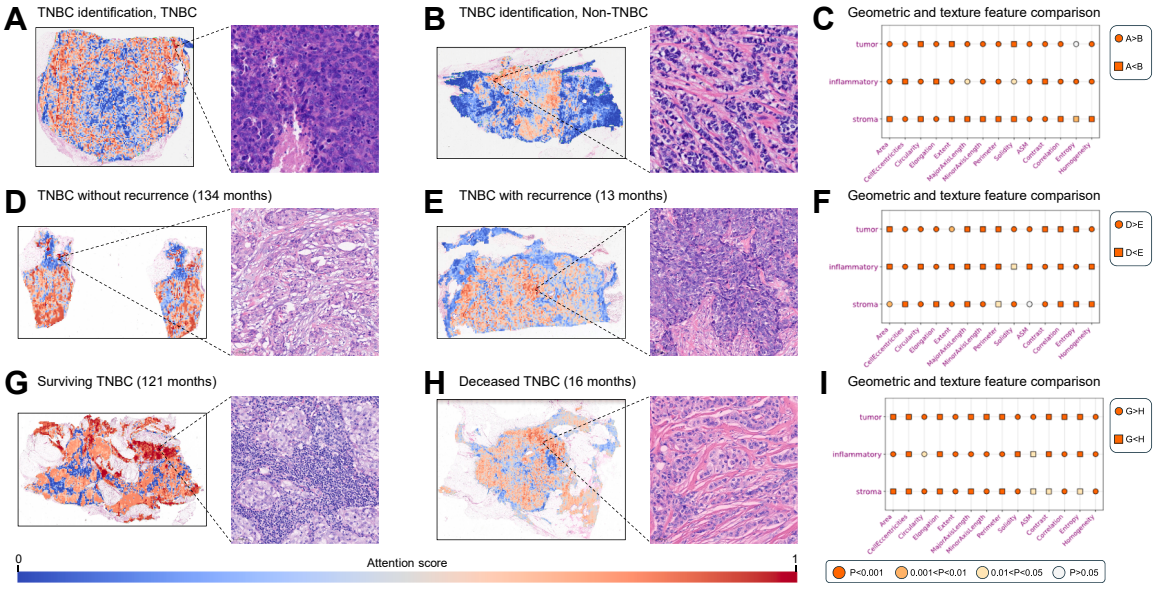

Fig. 4: **Heatmaps of our TNBC identification and prognosis prediction (TRIP) system on pathology images**. The top row shows the attention maps of triple-negative breast cancer (TNBC) identification model in TRIP system. We find that TNBC tumour cells exhibit obvious nuclear atypia with pleomorphic, irregular nuclei containing vesicular chromatin and prominent nucleoli, while central tumour necrosis is commonly present (A), whereas the nuclear atypia of tumour cells is not evident, nuclei is related regular, and tumour necrosis is rare in non-TNBC specimen (B). The second row exhibits the attention maps of disease-free survival prediction model in TRIP system. Both two samples have obvious tumour cell anomalies and rare interstitial lymphocytes and plasma cells, while mitotic activity in the sample with recurrence (E) is significantly more than sample with no recurrence (D). The third row illustrates the attention maps of overall survival prediction model in TRIP system. It can be observed that large lymphocytic and plasma cell infiltration in the interstitium in TNBC with good prognosis (G), while a lack of such patterns indicates poor prognosis (H). In subfigures C, F, and I, we show the comparison results of geometric and texture features between two samples on the left.

labour-intensive biomarker annotations,[22–27] TRIP introduces several key innovations. First, it employs a general-purpose pathology foundation model (GPFM)[31] to extract robust patch embeddings across staining variations. Second, it integrates a bidirectional Mamba-based encoder[42] module to model long-range spatial dependencies and an attention-based aggregation to effectively capture global contextual features in WSIs, a novel design beyond conventional MIL pooling or attention mechanisms.[30] Third, it employs test-time adaptation (TTA)[32] to dynamically adjust to distribution shifts across cohorts, enhancing robustness and generalisation during deployment. Collectively, these innovations enable TRIP to directly learn TNBC-specific morphological and prognostic features in a fully automated and scalable manner. Furthermore, the retrospective WSI data were collected independently at each clinical centre according to their routine diagnostic workflows without centralised protocol enforcement. As H&E staining variability across centres may introduce distributional shifts that affect model performance, the general pathology foundation model and TTA strategy adopted in our system can alleviate this issue. First, the pathology foundation models used in our pipeline (e.g., GPFM) were pretrained on tens of thousands of WSIs collected from diverse sources, encompassing a wide range of H&E staining protocols, scanners, and institutions. As a result, the extracted patch embeddings are inherently robust to common staining variations, reducing the need for additional stain normalisation procedures. Second, the TTA strategy fine-tunes the normalisation layers in the TRIP system during inference, while keeping all other model parameters fixed. As shown in our validation experiments, this adaptation method significantly improves model robustness across external cohorts, including those with distinct staining and acquisition characteristics.

From a clinical workflow perspective, TRIP can be flexibly deployed as (1) a triage tool to prioritise cases for confirmatory IHC/FISH testing, (2) a decision-support aid to validate ambiguous receptor status, or (3) a standalone diagnostic system in resource-constrained settings. In particular, TRIP offers a viable alternative in cases where tissue quality or availability limits molecular testing, such as poorly fixed specimens or emergency diagnoses. Moreover, the general pathology foundation model and TTA increase TRIP's generalisability, mitigating performance drops from non-standardised data across centres seen in prior AI models. Furthermore, its interpretability is

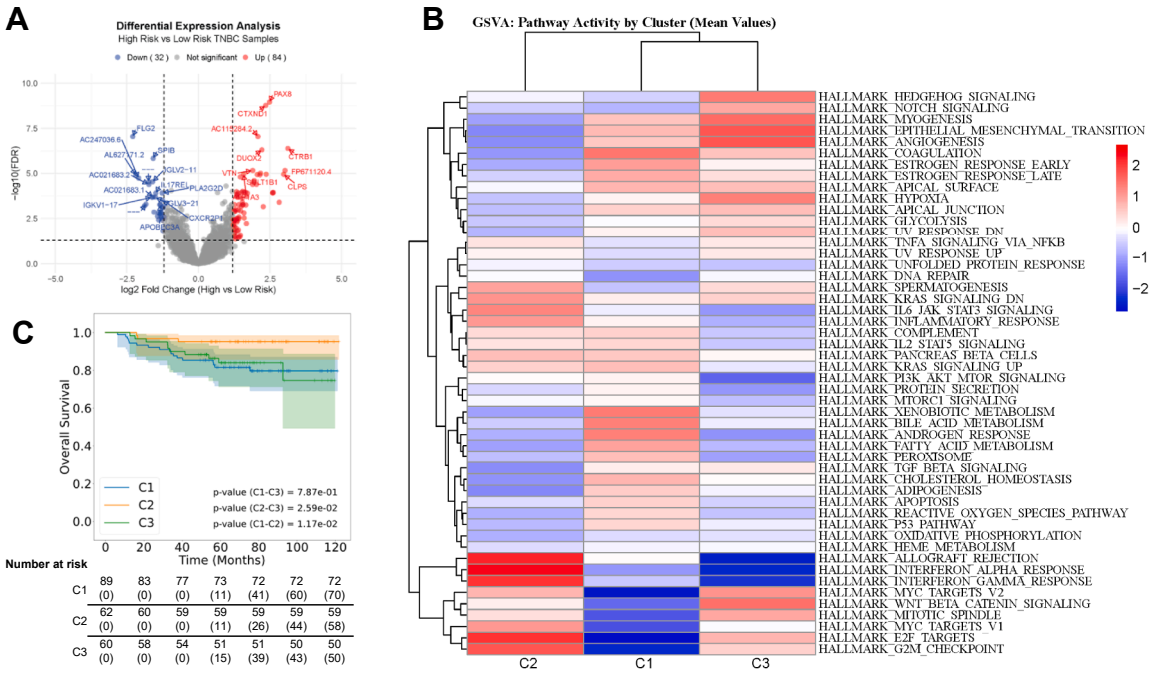

***Fig. 5*: Integrated multi-omics analysis of triple-negative breast cancer (TNBC) samples**. A. Differential gene expression analysis. Volcano plot comparing log2-fold changes (x-axis) versus -log 10(FDR-adjusted P-values) (y-axis) for genes differentially expressed between high-risk and low-risk TNBC samples predicted by TNBC identification and prognosis prediction (TRIP) system. Red dots indicate upregulated genes (n = 84), blue dots denote downregulated genes (n = 32), and grey dots represent non-significant genes. B. Gene set variation analysis (GSVA) pathway activity clustering. Heatmap of mean GSVA scores for Hallmark pathways across risk clusters (C1–C3). Colour changes from blue (low activity) to red (high activity) reflect pathway activation patterns. C. Kaplan–Meier analysis for C1–C3. C2 showcases significantly better prognosis than C1 and C3. The number of censored individuals in each group are in brackets.

enhanced via attention heatmaps, which highlight histopathological regions consistent with known TNBC phenotypes, such as desmoplastic stroma and high mitotic activity.[37,38] These model-localised visual cues increase trustworthiness and clinical transparency. In addition, TRIP is lightweight enough to be deployed on a workstation equipped with a single GPU (e.g., 12 GB memory), making it feasible for real-world implementation in most pathology departments.

Importantly, TRIP's prognostic predictions were also biologically grounded. Our integrated multi-omics analyses revealed three risk clusters with distinct immune and tumour-promoting signatures[40,41]—corroborating TRIP's stratifications with molecular pathways. In additional subgroup analyses, TRIP maintained predictive consistency across different age groups and HER2-low expression status, indicating resilience to molecular and demographic heterogeneity. Although BRCA1/2 mutation status was not available in our dataset, future efforts will focus on integrating this genomic information for more granular risk modelling.

Several recent studies have explored the use of WSI-based deep learning to predict treatment response in breast cancer.[43–47] These models typically leverage nuclear morphology, spatial TIL architecture, or multi-stain features to predict pathological complete response (pCR) after neoadjuvant chemotherapy. While promising, many of these approaches require labour-intensive annotations, specialised stains, or multi-modal inputs, which limit their scalability and real-world adoption. In contrast, TRIP operates directly on routine H&E slides without additional preprocessing, annotations, or molecular inputs. Its prognostic predictions are derived from a single diagnostic image, offering a scalable and interpretable alternative. As such, TRIP could support downstream treatment stratification and potentially guide de-escalation strategies in low-risk patients with TNBC, particularly in settings where complex multi-stage pipelines are infeasible.

This study has several limitations. First, our analysis was limited to post-resection specimens, excluding core needle biopsy (CNB) samples that are more common in early-stage diagnostic workflows. While TRIP's architecture is compatible with CNB, future validation is necessary to assess its efficacy in this setting. Second, clinical variables such as age, TNM stage, and Ki-67 were not incorporated into the model. Although TRIP was designed as an image-only system to maximise generalisability, incorporating multimodal data may

further enhance prognostic precision.[36] However, due to limited availability of complete clinical annotations across patients with TNBC, such integration remains challenging at present. Third, although we constructed the EXSURV cohort by aggregating small external validation cohorts to enhance statistical power, residual inter-cohort heterogeneity—such as variations in staining protocols and scanner devices—remains a potential confounding factor. While our model employs test-time adaptation to mitigate such distribution shifts, it may not fully eliminate centre-specific biases. Lastly, TRIP's clinical impact, particularly in guiding treatment de-escalation strategies, needs prospective validation. Preliminary results in untreated subgroups suggest that TRIP can identify low-risk patients who may safely omit adjuvant therapies, but further studies are warranted.

Future directions for improving this work include (1) extending TRIP's applicability to biopsy settings, (2) incorporating clinical and genomic data into a multimodal framework, and (3) initiating prospective validation to assess its real-world clinical utility. With its scalability, interpretability, and strong performance, TRIP has the potential to assist pathologists in identifying aggressive TNBC cases and tailoring risk-adapted treatment strategies.

Contributors

XMZ and HC initiated the project, organised a collaborative team, and provided study supervision. XMZ, HJZ, QC, XW, YJF, YLX, JZ, and HC conceived the study and its design. XMZ, QC, JPW, QYC, JLW, and JZ collected the data for the model's training. HJZ, XW, CJ, FTZ, and HC developed the network architectures, training, and validation setup. YJF, BL, MTH, CYY, XY, and YLX collected the data for the model's external validation. XMZ, HJZ, QC, XW, and YJF co-wrote the manuscript. YLX, JZ, and HC critically revised the manuscript, and all authors discussed the results and provided feedback regarding the manuscript. All authors had full access to the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. HJZ, XW, and HC have access to and verify the underlying study data. XMZ and HC had final responsibility for the decision to submit for publication.

Data sharing statement

The codes for the proposed TRIP model in this study have been deposited on Github (https://github.com/moothes/TRIP), which can only be used for non-commercial research purposes. The private histopathology images and mRNA expression data used in this study are not publicly available due to restrictions of hospital regulations and patient privacy. Specific conditions and restrictions of access to the datasets are to be discussed directly with the main investigators: XZ and QC from the First Affiliated Hospital of Zhejiang University. Public data from The Cancer Genome Atlas (TCGA) program were all obtained from the National Cancer Institute GDC Data Portal: https://portal.gdc.cancer.gov.

Declaration of interests

We declare no competing interests.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (No. 62202403 and No. 62402458), and Hong Kong Innovation and Technology Commission (Project No. MHP/002/22 and ITCPD/17–9), the Huadong Medicine Joint Fund of the Zhejiang Provincial Natural Science Foundation of China under Grant No. LHDMZ25H160002, and the Zhejiang Province Health Major Science and Technology Program of National Health Commission Scientific Research Fund (No. WKJ-ZJ-2426). We sincerely thank the collaborating hospitals for providing the histopathology images and clinical records of patients with breast cancer. Moreover, we gratefully acknowledge Xiao-Ke Ma and Min Zhang (Xidian University) and Ming Yang (The Hong Kong University of Science and Technology) for their valuable suggestions in integrated multi-omics analysis.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2025.103557.

References

1 Dent R, Trudeau M, Pritchard KI, et al. Triple-negative breast cancer: clinical features and patterns of recurrence. *Clin Cancer Res*. 2007;13(15):4429–4434.
2 Thagaard J, Stovgaard ES, Vognsen LG, et al. Automated quantification of sTIL density with H&E-based digital image analysis has prognostic potential in triple-negative breast cancers. *Cancers*. 2021;13(12):3050.
3 Bianchini G, Balko JM, Mayer IA, Sanders ME, Gianni L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nat Rev Clin Oncol*. 2016;13(11):674–690.
4 Rakha EA, El-Sayed ME, Green AR, Lee AH, Robertson JF, Ellis IO. Prognostic markers in triple-negative breast cancer. *Cancer*. 2007;109(1):25–32.
5 Cleator S, Heller W, Coombes RC. Triple-negative breast cancer: therapeutic options. *Lancet Oncol*. 2007;8(3):235–244.
6 Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S. Pathologists' guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *Breast Care*. 2010;5(3):185–187.
7 Albusayli R, Graham JD, Pathmanathan N, et al. Artificial intelligence-based digital scores of stromal tumour-infiltrating lymphocytes and tumour-associated stroma predict disease-specific survival in triple-negative breast cancer. *J Pathol*. 2023;260(1):32–42.
8 Savas P, Salgado R, Denkert C, et al. Clinical relevance of host immunity in breast cancer: from TILs to the clinic. *Nat Rev Clin Oncol*. 2016;13(4):228–241.
9 Adams S, Gray RJ, Demaria S, et al. Prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancers from two phase III randomized adjuvant breast cancer trials: ECOG 2197 and ECOG 1199. *J Clin Oncol*. 2014;32(27):2959–2966.
10 Chen RJ, Lu MY, Williamson DF, et al. Pan-cancer integrative histology-genomic analysis via multimodal deep learning. *Cancer Cell*. 2022;40(8):865–878.
11 Luo L, Wang X, Lin Y, et al. Deep learning in breast cancer imaging: a decade of progress and future directions. *IEEE Rev Biomed Eng*. 2024;18:130–151.
12 Xu Y, Chen H. *Multimodal optimal transport-based co-attention transformer with global structure consistency for survival prediction*. IEEE/CVF International Conference on Computer Vision; 2023:21241–21251.
13 Zhou F, Chen H. *Cross-modal translation and alignment for survival analysis*. IEEE/CVF International Conference on Computer Vision; 2023:21485–21494.
14 Wang X, Chen H, Gan C, et al. Weakly supervised deep learning for whole slide lung cancer image analysis. *IEEE Trans Cybern*. 2019;50(9):3950–3962.
15 Ilse M, Tomczak J, Welling M. *Attention-based deep multiple instance learning. International conference on machine learning*. 2018:2127–2136.
16 Yang S, Wang Y, Chen H. *Mambamil: enhancing long sequence modeling with sequence reordering in computational pathology. International conference on medical image computing and computer-assisted intervention*. 2024:296–306.
17 Amgad M, Hodge JM, Elsebaie MA, et al. A population-level digital histologic biomarker for enhanced prognosis of invasive breast cancer. *Nat Med*. 2024;30(1):85–97.
18 Ahn B, Moon D, Kim HS, et al. Histopathologic image–based deep learning classifier for predicting platinum-based treatment responses in high-grade serous ovarian cancer. *Nat Commun*. 2024;15(1):4253.

19 Gao Y, Ventura-Diaz S, Wang X, et al. An explainable longitudinal multi-modal fusion model for predicting neoadjuvant therapy response in women with breast cancer. *Nat Commun*. 2024;15(1):9613.
20 Iyer JS, Juyal D, Le Q, et al. AI-based automation of enrollment criteria and endpoint assessment in clinical trials in liver diseases. *Nat Med*. 2024;30(10):2914–2923.
21 Tsai PC, Lee TH, Kuo KC, et al. Histopathology images predict multi-omics aberrations and prognoses in colorectal cancer patients. *Nat Commun*. 2023;14(1):1–3.
22 Sun P, He J, Chao X, et al. A computational tumor-infiltrating lymphocyte assessment method comparable with visual reporting guidelines for triple-negative breast cancer. *eBioMedicine*. 2021;70.
23 Hou X, Li X, Han Y, et al. Triple-negative breast cancer survival prediction using artificial intelligence through integrated analysis of tertiary lymphoid structures and tumor budding. *Cancer*. 2024;130(S8):1499–1512.
24 Yuan Y, Failmezger H, Rueda OM, et al. Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Sci Transl Med*. 2012;4(157):157ra143.
25 Lu Z, Xu S, Shao W, et al. Deep-learning–based characterization of tumor-infiltrating lymphocytes in breast cancers from histopathology images and multiomics data. *JCO Clin Cancer Inform*. 2020;4:480–490.
26 Jang W, Lee J, Park KH, Kim A, Lee SH, Ahn S. Molecular classification of breast cancer using weakly supervised learning. *Cancer Res Treat*. 2024;57(1):116.
27 Zhao S, Yan CY, Lv H, et al. Deep learning framework for comprehensive molecular and prognostic stratifications of triple-negative breast cancer. *Fundam Res*. 2024;4(3):678–689.
28 Gamble P, Jaroensri R, Wang H, et al. Determining breast cancer biomarker status and associated morphological features using deep learning. *Commun Med*. 2021;1(1):14.
29 Lehmann BD, Colaprico A, Silva TC, et al. Multi-omics analysis identifies therapeutic vulnerabilities in triple-negative breast cancer subtypes. *Nat Commun*. 2021;12(1):6276.
30 Lu MY, Williamson DF, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng*. 2021;5(6):555–570.
31 Ma J, Guo Z, Zhou F, et al. Towards a generalizable pathology foundation model via unified knowledge distillation. *Nat Biomed Eng*. 2025. https://doi.org/10.1038/s41551-025-01488-4.
32 Wang D, Shelhamer E, Liu S, Olshausen B, Darrell T. *Tent: fully test-time adaptation by entropy minimization. International conference on learning representations*. 2021.
33 Zhao S, Chen DP, Fu T, et al. Single-cell morphological and topological atlas reveals the ecosystem diversity of human breast cancer. *Nat Commun*. 2023;14(1):6796.
34 Chen RJ, Ding T, Lu MY, et al. Towards a general-purpose foundation model for computational pathology. *Nat Med*. 2024;30(3):850–862.
35 Huang Z, Bianchi F, Yuksekgonul M, Montine TJ, Zou J. A visual–language foundation model for pathology image analysis using medical twitter. *Nat Med*. 2023;29(9):2307–2316.
36 Yang Z, Guo C, Li J, et al. An explainable multimodal artificial intelligence model integrating histopathological microenvironment and EHR phenotypes for germline genetic testing in breast cancer. *Adv Sci*. 2025;e02833.
37 Bonzanini M, Morelli L, Bonandini EM, Leonardi E, Pertile R, Dalla Palma P. Cytologic features of triple-negative breast carcinoma. *Cancer Cytopathol*. 2012;120(6):401–409.
38 Chivukula M, Striebel JM, Ersahin Ç, Dabbs DJ. Evaluation of morphologic features to identify "basal-like phenotype" on core needle biopsies of breast. *Appl Immunohistochem Mol Morphol*. 2008;16(5):411–416.
39 Lehmann BD, Jovanović B, Chen XI, et al. Refinement of triple-negative breast cancer molecular subtypes: implications for neoadjuvant chemotherapy selection. *PLoS One*. 2016;11(6):e0157368.
40 Sleeman JP, Christofori G, Fodde R, et al. Concepts of metastasis in flux: the stromal progression model. *Semin Cancer Biol*. 2012;22(3):174–186.
41 Oshi M, Gandhi S, Yan L, et al. Abundance of reactive oxygen species (ROS) is associated with tumor aggressiveness, immune response, and worse survival in breast cancer. *Breast Cancer Res Treat*. 2022;194(2):231–241.
42 Dao T, Gu A. *Transformers are SSMs: generalized models and efficient algorithms through structured state space duality. International conference on machine learning*. 2024:10041–10071.
43 Naylor P, Lazard T, Bataillon G, et al. Prediction of treatment response in triple negative breast cancer from whole slide images. *Front Signal Process*. 2022;2:851809.
44 Dodington DW, Lagree A, Tabbarah S, et al. Analysis of tumor nuclear features using artificial intelligence to predict response to neoadjuvant chemotherapy in high-risk breast cancer patients. *Breast Cancer Res Treat*. 2021;186:379–389.
45 Corredor G, Toro P, Lu C, et al. Computational features of tumor-infiltrating lymphocyte architecture of residual disease after chemotherapy on H&E images as prognostic of overall and disease-free survival for triple-negative breast cancer. *J Clin Oncol*. 2021;39:584.
46 Ogier du Terrail J, Leopold A, Joly C, et al. Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer. *Nat Med*. 2023;29(1):135–146.
47 Duanmu H, Bhattarai S, Li H, et al. *Spatial attention-based deep learning system for breast cancer pathological complete response prediction with serial histopathology images in multiple stains. International conference on medical image computing and computer-assisted intervention*. 2021:550–560.