Journal of Pathology
J Pathol 2023
Published online 24 February 2023 in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/path.6061

ORIGINAL ARTICLE

Artificial intelligence-based digital scores of stromal tumour-infiltrating lymphocytes and tumour-associated stroma predict disease-specific survival in triple-negative breast cancer

Rawan Albusayli[1], J Dinny Graham[2,3], Nirmala Pathmanathan[3,4], Muhammad Shaban[5], Shan E Ahmed Raza[1], Fayyaz Minhas[1], Jane E Armes[6] and Nasir Rajpoot[1,7]*

[1] *Tissue Image Analytics Centre, The University of Warwick, Coventry, UK*
[2] *The Westmead Institute for Medical Research, The University of Sydney, Sydney, NSW, Australia*
[3] *Westmead Breast Cancer Institute, Western Sydney Local Health District, Sydney, NSW, Australia*
[4] *Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia*
[5] *Harvard Medical School, Harvard University, Boston, MA, USA*
[6] *Pathology Queensland, Queensland Health, Herston, QLD, Australia*
[7] *The Alan Turing Institute, London, UK*

Correspondence to: N Rajpoot, Tissue Image Analytics (TIA) Centre, Department of Computer Science, The University of Warwick, Coventry CV4 7AL, UK. E-mail: n.m.rajpoot@warwick.ac.uk

Abstract

Triple-negative breast cancer (TNBC) is known to have a relatively poor outcome with variable prognoses, raising the need for more informative risk stratification. We investigated a set of digital, artificial intelligence (AI)-based spatial tumour microenvironment (sTME) features and explored their prognostic value in TNBC. After performing tissue classification on digitised haematoxylin and eosin (H&E) slides of TNBC cases, we employed a deep learning-based algorithm to segment tissue regions into tumour, stroma, and lymphocytes in order to compute quantitative features concerning the spatial relationship of tumour with lymphocytes and stroma. The prognostic value of the digital features was explored using survival analysis with Cox proportional hazard models in a cross-validation setting on two independent international multi-centric TNBC cohorts: The Australian Breast Cancer Tissue Bank (AUBC) cohort ($n = 318$) and The Cancer Genome Atlas Breast Cancer (TCGA) cohort ($n = 111$). The proposed digital stromal tumour-infiltrating lymphocytes (Digi-sTILs) score and the digital tumour-associated stroma (Digi-TAS) score were found to carry strong prognostic value for disease-specific survival, with the Digi-sTILs and Digi-TAS scores giving C-index values of 0.65 ($p = 0.0189$) and 0.60 ($p = 0.0437$), respectively, on the TCGA cohort as a validation set. Combining the Digi-sTILs feature with the patient's positivity status for axillary lymph nodes yielded a C-index of 0.76 on unseen validation cohorts. We surmise that the proposed digital features could potentially be used for better risk stratification and management of TNBC patients.

Keywords: triple-negative breast cancer; tumour microenvironment; disease-specific survival analysis; histology images; artificial intelligence

Received 10 August 2022; Revised 12 January 2023; Accepted 23 January 2023

Conflict of interest statement: NR is CSO and a co-founder of Histofy Ltd. No other conflicts of interest were declared.

Introduction

Triple-negative breast cancer (TNBC) is a subtype of breast cancer defined by its lack of expression of oestrogen receptor (ER) and progesterone receptor (PR), and absence of overexpression of the human epidermal growth factor receptor 2 (HER2). In the United States, TNBC accounts for about 15–20% of breast cancer cases and is more frequent among African American and Hispanic women who are young and have a relatively high pre-menopausal body mass index [1–3]. The disease mainly occurs in young women and confers a shorter survival time and a 40% mortality rate within the first 5 years after diagnosis, and 75% within 3 months after recurrence [4]. The median survival time after metastasis is 13.3 months, because 46% of TNBC patients develop distant metastasis that often involves the brain and visceral organs [5].

Chemotherapy is the main treatment option as TNBC patients do not benefit from targeted therapies, due to the

absence of three receptors, and no molecular targets have been recognised for the subtype. However, TNBC patients' response to chemotherapy lacks durability, with high relapse rates [6]. Different potential treatments have been proposed in recent years to improve the disease outcomes, such as optimising chemotherapy regimens and incorporating immunotherapy [7,8]. Understanding the disease microenvironment is crucial to giving potential reasons for treatment responses and predicting markers for favourable survival outcomes.

Recently, the immunogenic microenvironment and tumour–stroma interactions have been found to be of clinical significance [9–11]. The presence of tumour-infiltrating lymphocytes (TILs) in the breast tumour microenvironment increased pathological complete responses (pCR) with promising results for immune checkpoint inhibitors in a subgroup of metastatic TNBCs [6,12,13]. Early-stage TNBC patient outcomes correlate strongly with the presence of stromal TILs (sTILs), as an 18% death risk reduction was associated with a 10% increment of stromal lymphocytes [14]. Many studies conclude that a higher number of sTILs, measured using a unified methodology, show a favourable response to chemotherapy and neoadjuvant therapy and can be used as a predictive marker in TNBC [6,15–17]. However, due to the inherent subjective nature of manual assessment, this feature is not reported in a routine manner. The use of more accurate and systematic digital methods for the measurement of this characteristic presents the opportunity to add clinically relevant information that complements traditional assessment approaches.

The stroma surrounding tumour tissue carries a prognostic value and impacts tumour development and behaviour in triple-negative cancer subpopulations [18]. Patients with stroma-high tumours show a worse prognosis compared with women with a low tumour–stroma ratio. The tumour–stroma ratio shows discriminative results in triple-negative tumours compared with non-TNBC despite the patient's clinical variables such as age and tumour size [11]. The clinical risk stratification between the two groups was based on evaluating digital haematoxylin and eosin (H&E)-stained histological sections with some variation between methodologies [11,19,20].

In general, TNBC has a high histological grade (grade III), greater mean tumour size, and elevated mitotic count. Other features such as central necrosis, pushing margins of invasion, a lymphocytic stromal response, and multiple apoptotic cells also are common [3,5]. However, despite all the recent progress to understand the disease, it is widely accepted that TNBC cases are grouped according to what they lack rather than what they share, and they are posing a significant clinical management challenge [1,3,10].

Currently, histopathological analysis remains the mainstay of TNBC prognostication. However, the digitisation of tissue slides into whole-slide images (WSIs) opens the possibility of studying the spatial tumour microenvironment (sTME) to mine for novel digital image-based markers using modern artificial intelligence (AI) and deep learning-based image analysis [21]. Using these technologies, it is now possible to explore quantitative measures of the TNBC sTME instead of relying only on qualitative assessment, allowing an in-depth investigation of the TNBC sTME by expanding the observation space.

Deep convolutional neural networks (CNNs) have emerged as the leading branch of AI for image analysis, demonstrating significant strength in several computer vision and pattern recognition tasks [22,23]. The deep CNNs have been shown to effectively classify and segment tissue types and different kinds of nuclei in the WSIs with a high degree of accuracy, opening vast opportunities for understanding the sTME from novel perspectives [15,24–27]. Several researchers have recently proposed computational models to assess the role of different biomarkers such as the absolute mitotic count [28], tumour-infiltrating lymphocyte architecture [29], and spatial organisation of TILs [30].

This study combines the power of deep CNNs and graph theory to develop novel digitised sTME markers for TNBC prognostication. After detailed analysis, we found that digital markers of the tumour-associated stroma (TAS) and stromal tumour-infiltrating lymphocytes (sTILS) could stratify TNBC cases from two independent multi-centric cohorts (one from Australia and the other from the USA) into two risk groups with statistically significant differences in terms of disease-specific survival. Patient survival predictions can be applied to tailored clinical therapy and expand response assessment aspects in future clinical trials.

Materials and methods

Study cohort and datasets

In this study, we analysed data from two independent multi-centric TNBC datasets, consisting of digitised whole-slide images (WSIs) of H&E-stained slides with linked clinical follow-up and outcome data, to develop and evaluate quantitative measures derived from our deep learning-based analysis of the WSIs. The two datasets typically contain a single WSI per patient. The two cohorts are denoted as the AUBC (Australian Breast Cancer) cohort ($n = 318$) and the TCGA (The Cancer Genome Atlas) cohort ($n = 111$). The use of the AUBC cohort was approved by the Western Sydney Local Health District Human Research Ethics Committee (approval reference 2019/ETH02413) and the Royal Brisbane and Women's Hospital Human Research Ethics Committee (HREC/2021/QRBW/76049). The AUBC images were obtained from the Australian Breast Cancer Tissue Bank, and donors provided informed consent to the bank for their tissue and data to be used for future unspecified research at the time of recruitment. The TCGA cohort was de-identified and publicly available, and participant consent had previously been obtained by The Cancer Genome Atlas Consortium.

Table 1. Summary of clinical and pathological features for the patients in the two cohorts.

Variable	Sub-variables	AUBC *n* (%) or mean (SD)	TCGA *n* (%) or mean (SD)
Age (months)	0–188	56 (14.89)	55 (11.94)
Laterality	Bilateral	3 (00.94%)	
	Left	164 (51.57%)	63 (56.75%)
	Right	151 (47.48%)	48 (43.24%)
Primary histological diagnosis	Invasive lobular carcinoma (ILC)	6 (01.88%)	1 (0.9%)
	No special type (NST)	283 (88.99%)	101 (99.99%)
	Medullary cancer	6 (01.88%)	2 (1.80%)
	Metaplastic cancer	6 (01.88%)	1 (0.9%)
	Other	17 (05.34%)	6 (5.40%)
Tumour stage	T1	98 (30.81%)	NR
	T2	196 (61.64)	
	T3	21 (6.61%)	
	NA	3 (0.94%)	
Grade	2	19 (05.97%)	NR
	3	295 (92.76%)	
	No grade	4 (01.25%)	
Stage	I	NR	21 (18.91%)
	II		73 (65.76%)
	III		13 (11.71%)
	Others		4 (03.60%)
Axillary lymph nodes positive	0	220 (69.18%)	78 (70.27%)
	>0	98 (30.81%)	33 (29.72%)

N, number of patients; NA, not available; NR, data not reported; SD, standard deviation.

The AUBC dataset collected by the Australian Breast Cancer Tissue Bank consisted of 412 cases, with one H&E-stained representative slide per case, treated primarily with surgery between 2003 and 2017, from 12 different pathology departments. All digitised WSIs for the slides were scanned using a Hamamatsu NanoZoomer scanner (Hamamatsu Photonics, Bridgewater, NJ, USA) with a magnification power of 40×. The WSIs were further filtered to exclude cases with missing follow-up data and inadequate imaging, leaving a total of 318 cases in the AUBC cohort for the study (supplementary material, Figure S1).

The TCGA dataset consisted of WSIs for the diagnostic slides, typically one slide per case, and the corresponding survival data for a total of 157 TNBC cases. They were retrieved from the TCGA portal from a larger collection of 1,098 breast cancer cases based on their negativity for the three receptors (ER, PR, and HER2), with their initial pathological diagnosis year varying between 1993 and 2013. After performing quality checks, 46 slides with artefacts were excluded, resulting in 111 cases (111 WSIs) for the TCGA-TNBC cohort (supplementary material, Table S1), referred to as the TCGA cohort hereafter.

Clinicopathological features

In total, 429 female patients were included in the study, with the median follow-up duration of the whole cohort being 45 months (range 0–188 months) (Table 1). The follow-up duration for the AUBC cohort was up to 188 months compared with that of the TCGA cohort, which had a maximum follow-up duration of 133 months. To harmonise the follow-up duration between the two cohorts, we right-censored the maximum survival analysis time to 120 months (10 years). The average age of all participants was 56 years, and the disease caused death for 24% of the patients. The majority of the TNBC lesions were grade III (92.76% of AUBC), and no prognostic factors were found in the disease-specific survival (DSS) analysis using patients' grade (supplementary material, Figure S2).

Tissue detection and segmentation

A concept diagram of our analytical pipeline is shown in Figure 1. After dividing the digitised WSIs into smaller manageable image patches of size 224 × 224 pixels at 40× magnification power, the first step in our pipeline was to recognise and localise the various tissue types present in the WSIs. We employed a previously developed in-house deep learning model [31] to automatically recognise the different tissue types, followed by a graph-based method to enhance the class predictions based on neighbouring nodes.

We used an end-to-end CNN model to label a given image patch into one of the six classes: stroma, lymphocytes, necrosis, tumour, normal lobules, and other (including miscellaneous other tissue types and the fatty tissue), employing the ResNet neural network model [32] to perform initial classification of the tissue types. An expert pathologist (JA) manually annotated regions corresponding to the six tissue types (stroma, lymphocytes, necrosis, tumour, normal lobules, and others) using the Automated Slide Analysis Platform (ASAP) annotation tool [33] for training of the tissue classification algorithm. The batch size was set to 16 and the network was trained for 100 epochs with a learning rate of 0.001. The data were split into an average of 80% and 20% for training and testing, respectively; no patches from the same slide were simultaneously used for discovery and validation (supplementary material,

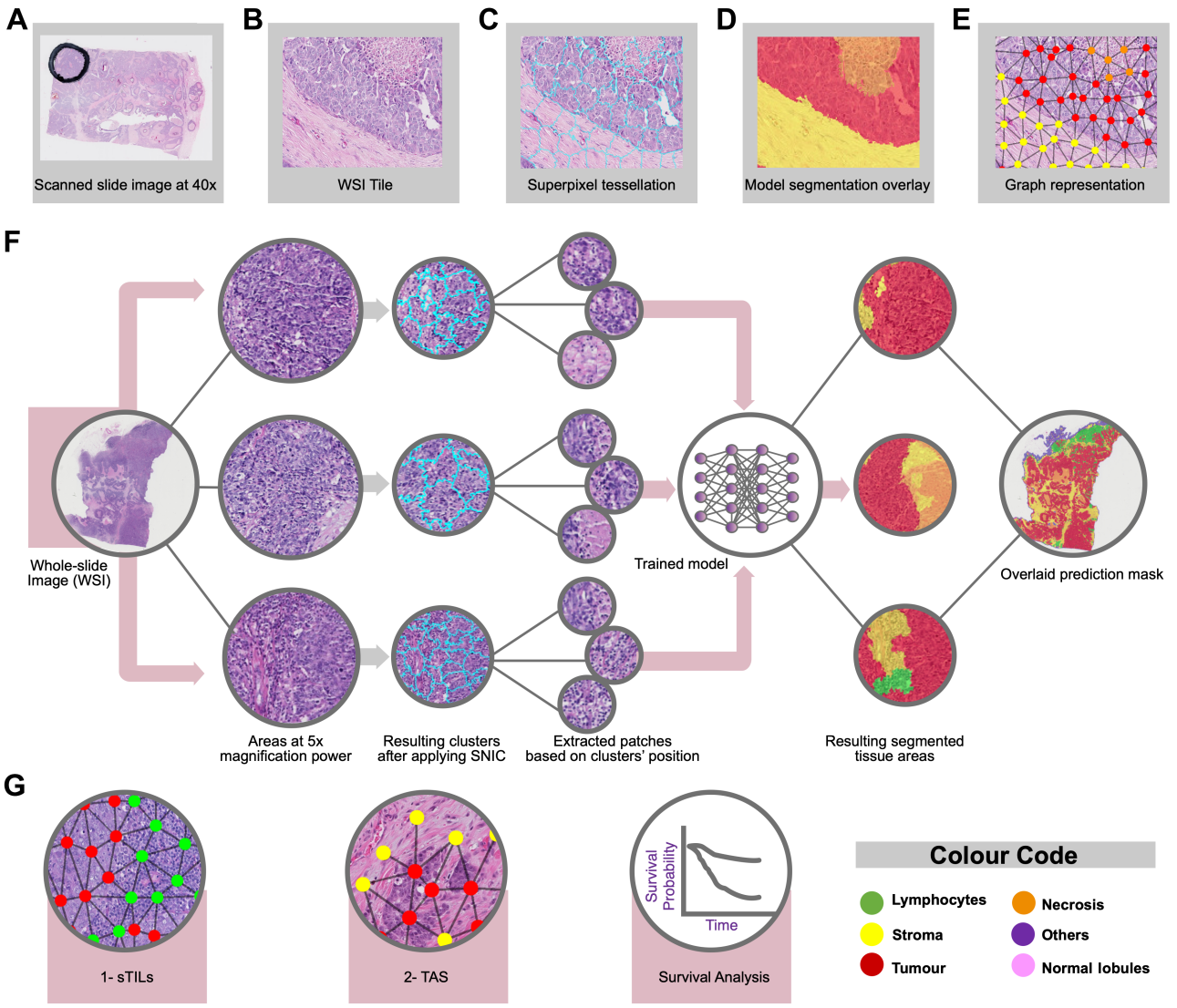

Figure 1. An overview of our analysis pipeline with visual representations of features.

Table S2), with the validation set's model performance having an average F1 score of 0.92 (supplementary material, Table S3).

At the time of inference on WSIs, we mapped the model prediction to patches after extracting superpixel clusters using the Simple Non-Iterative Clustering (SNIC) method [34] on image tiles at 5× magnification. The SNIC algorithm was applied to each WSI to segment tissue regions (superpixel), which represent patches, with an additional step to improve tissue prediction performance using a graph-based method to incorporate the effect of relationships between neighbouring nodes [31].

Extraction of spatial tumour microenvironment (sTME) features

The graph representation of each WSI was used to extract quantitative sTME image features for each patient. The extracted features included metrics derived from the connectivity of graph nodes of different tissue types. Supplementary material, Table S4 gives the equations and descriptions for each of the selected features.

The sTILs feature was studied by finding associations between lymphocytes and stroma, as shown in Figure 1G, and the calculation of the Digi-sTILs score is based on the analysis of the entire WSI (500 patches per WSI on average). It shows strong risk stratification, which makes it a potential digital marker for extracting TNBC case outcomes. Tumour-associated stroma (TAS) density was investigated by finding and quantifying direct connections between the tumour and stroma nodes. The absence of such links excludes nodes from being represented in the corresponding features. Examples of these nodes' connectivity are shown in Figure 2G. The area selected for the quantification method is based on the recommendation from a clinical perspective [20], with one representative slide being used per case and WSI's highest Digi-TAS score patch considered for survival analysis. Different quantitative

Table 2. Univariate disease-specific survival analysis and survival machine learning model (SurvivalSVM) were performed on the two cohorts using the studied features.

Validation cohort	TCGA			AUBC		
Features	C-index	HR	*P* value	C-index	HR	*P* value
Digi-sTILs score	0.65	3.8	**0.0189**	0.60	2.5	**0.0054**
Digi-TAS score	0.60	3.2	**0.0437**	0.59	2.6	**0.0008**
Digi-sTILs + ALN (SurvivalSVM)	0.76	4.5	**0.0131**	0.71	4.0	**<0.0001**

The table shows the results of using each cohort as independent validation sets, with detailed results of the training sets that can be found in supplementary material, Table S5. The *p* values shown in bold font represent significant differences in the corresponding K–M curves.

Table 3. Univariate disease-specific survival analysis and survival machine learning model (SurvivalSVM) were performed with internal cross-validation in the AUBC cohort and mixed cohorts (AUBC, TCGA).

Validation cohorts	AUBC (two-folds, ten runs)			Mixed cohorts (two-folds, ten runs)		
Features	C-index (SD)	HR	*P* value	C-index (SD)	HR	*P* value
Digi-sTILs score	0.63 (0.03)	3.4	**0.0037**	0.61 (0.02)	2.8	**0.0103**
Digi-TAS score	0.58 (0.03)	5.3	**0.0044**	0.57 (0.02)	5.9	**<0.0001**
Digi-sTILs + ALN (SurvivalSVM)	0.73 (0.04)	6.9	**<0.0001**	0.68 (0.03)	4.7	**<0.0001**

The table shows the results of the validation sets, with detailed results of training sets that can be found in supplementary material, Tables S6 and S7. HR represents the hazard ratio, and SD denotes the standard deviation across the ten runs. The *p* values shown in bold font represent significant differences in the corresponding K–M curves.

combinations of tissue types were investigated using some or all tissue types; however, no significant results were found among them.

Survival analysis using sTME features

The follow-up time in months was defined as the time interval between the date of diagnosis and the date of death or the total duration of the follow-up. Patients who died from other causes, still alive without an event at the last reported date, or alive and remaining disease-free were labelled as censored. Statistical analysis was conducted using R software (version 3.6.3; https://www.r-project.org/) and Python 3.7 (Python Software Foundation, https://www.python.org/).

Two-sided statistical tests were performed with a significance level set at $p < 0.05$. Prognostic evaluation and survival analyses were performed using the Kaplan–Meier (KM) analysis to show a statistically significant different outcome between two groups of patients. The grouping was done using a cut-off on the feature value that corresponds to the most important relationship with the disease-specific survival of the discovery set based on the log-rank test with the same cut-off value used on the validation set. The feature scores below the feature cut-off correspond to the low-value group, while the high values represent the high-value group.

We report the concordance index (C-index), defined as the fraction of concordant pairs of individuals divided by the total number of all possible pairs based on Harrell's C-statistics [35,36]. In the case of survival analysis, the C-index assesses the model's discrimination power by measuring the correlation rank between the predicted risk scores and survival times [37]. A Cox proportional hazards (Cox PH) model was trained using either of the two cohorts as a discovery cohort to predict survival in the other cohort as the validation cohort. Based on the significance of the model's *p* value, we used the function *concordance.index* in the R survcomp package using the Cox PH model trained on the discovery set [38]. The C-indices for Survival Support Vector Machine (SurvivalSVM) were calculated based on the model's prediction risk score without splitting the patients into two groups, as with the KM analysis.

We expanded our analysis of time-to-event outcomes to use a combination of features with SurvivalSVM, which are both extensions of their standard models to right-censored time-to-event data [39]. We plotted the KM survival curves of the two risk groups with their C-index, *p*-value, and HR on validation sets. The main contribution of the SurvivalSVM model is that it finds the relationships between features and survival groups using a complex and nonlinear hyperplane. The high-dimensional feature space mapping gives a new direction to understand the association with time-to-event patient data. The split quality is measured using the log-rank test with a significant *p* value below 0.05. We used *scikit-survival*, which is a Python module for survival analysis (https://scikit-survival.readthedocs.io/) with their default hyperparameters. Kaplan–Meier survival curves were drawn for each experiment, and Cox proportional hazards regression was applied to determine C-indices, HRs, and *p* values.

Results

Experiments for the aforementioned features were conducted in four different ways. For independent external cross-validation, we used AUBC as the discovery

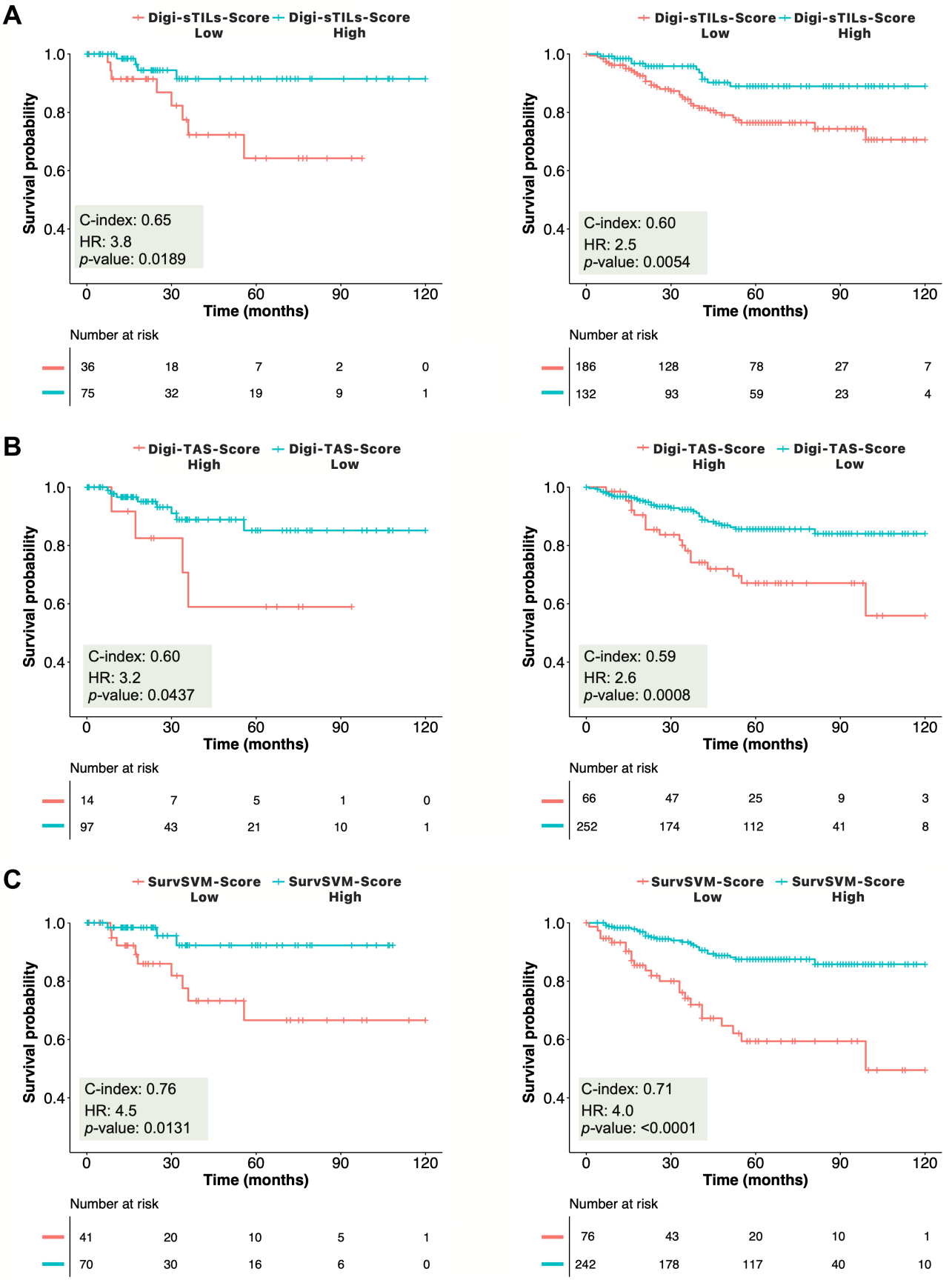

Figure 2. Significant survival differences for TCGA and AUBC. Kaplan–Meier (KM) curves are shown for high-value and low-value groups for the specific features when using a specific cohort for validation.

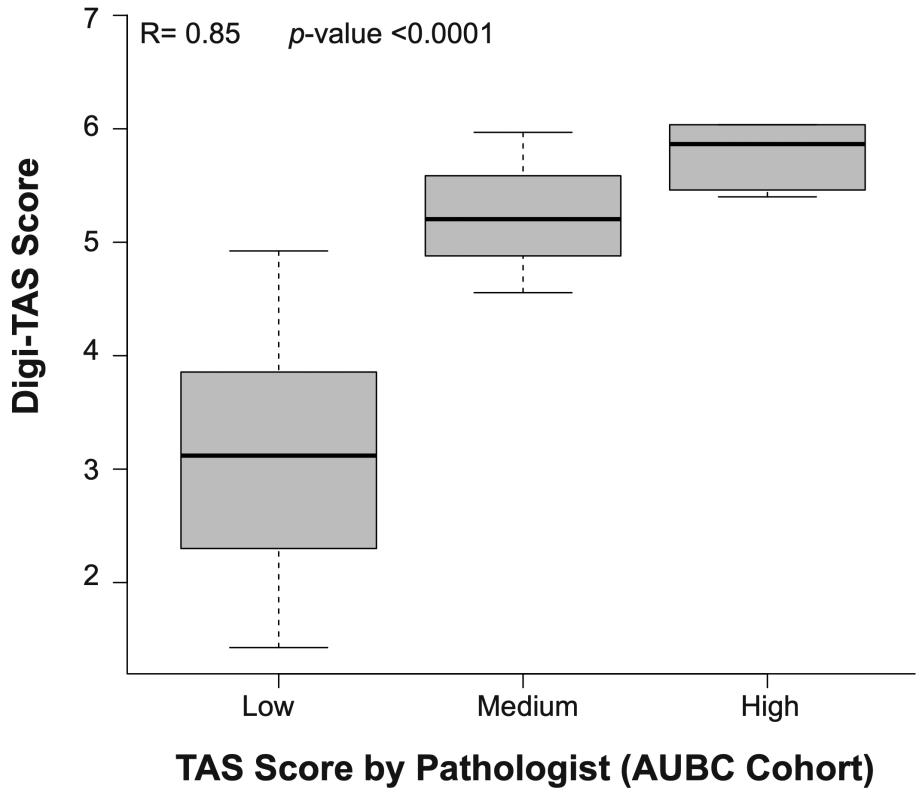

Figure 3. Illustrating the Spearman correlation with pathologists' TAS score and our digital scores.

cohort and TCGA as the external validation cohort and vice versa. We also performed mixed cohorts and AUBC internal cross-validation with two-fold cross-validation and ten multiple runs. For mixed cohorts and AUBC internal cross-validation, we report twice the median of the p values to avoid the inherent problems associated with data splitting and dependent p values [40].

Prognostic value of the digital score of TAS and sTILs scores

We conducted univariate Cox PH regression on each of the computed features and identified two significant features (log-rank $p < 0.05$) associated with survival and compatible with clinical viewpoints (Tables 2 and 3). The survival analysis was performed with KM estimates for low and high mortality risk groups based on the patient prediction scores for each feature (Figure 2).

Figure 2A illustrates the KM plots for TCGA and AUBC as validation sets for the digital sTILs (Digi-sTILs) feature, with the cohort stratified according to digital sTILs value. A high Digi-sTILs score was significantly associated with higher breast cancer disease-specific survival. The log-rank p value for the TCGA cohort as validation set was 0.0189, and 0.0054 for the AUBC cohort as validation set. The C-indices were calculated using the Cox PH regression model with 0.65, 0.60 for TCGA and AUBC for external cross-validation, respectively, and 0.63, 0.61 for AUBC with internal cross-validation and mixed cohorts with hazard ratios (HRs) between 2.8 and 3.4 (Table 3). We also investigated the concordance between the pathologist-assigned sTILs percentages from three pathologists [41] and the Digi-sTILs scores in the TCGA cohort (supplementary material, Figure S3).

Figure 2B shows the KM curves for the digital TAS (Digi-TAS) score. The curves show statistically significant differences between the high and low value groups with respect to outcomes, with a high value associated with shorter survival (HRs 2.6–5.9) and the C-indices ranging between 0.57 and 0.60 (Tables 2 and 3). The pathologist assigned TAS ratios (high/medium/low) manually to 20 cases from the AUBC cohorts, giving a high value of the correlation score (0.85) and the corresponding p value less than 0.0001 with Digi-TAS scores (Figure 3).

Multivariate assessment with a survival machine learning-based model

We investigated the performance of survival analysis with the combined features using the machine learning-based SurvivalSVM algorithm. Overall, the SurvivalSVM model performed quite well, with C-indices reaching 0.76 (Tables 2 and 3). The best feature combination in terms of the C-index is a combination of the Digi-sTILs score and the patient's positivity status to the axillary lymph node (ALN).

For the TCGA dataset as the external validation cohort, we achieved a high C-index (0.76) ($p = 0.0131$ and HR of 4.5) using SurvivalSVM with a similar trend observed using AUBC as a validation set with the C-index value of

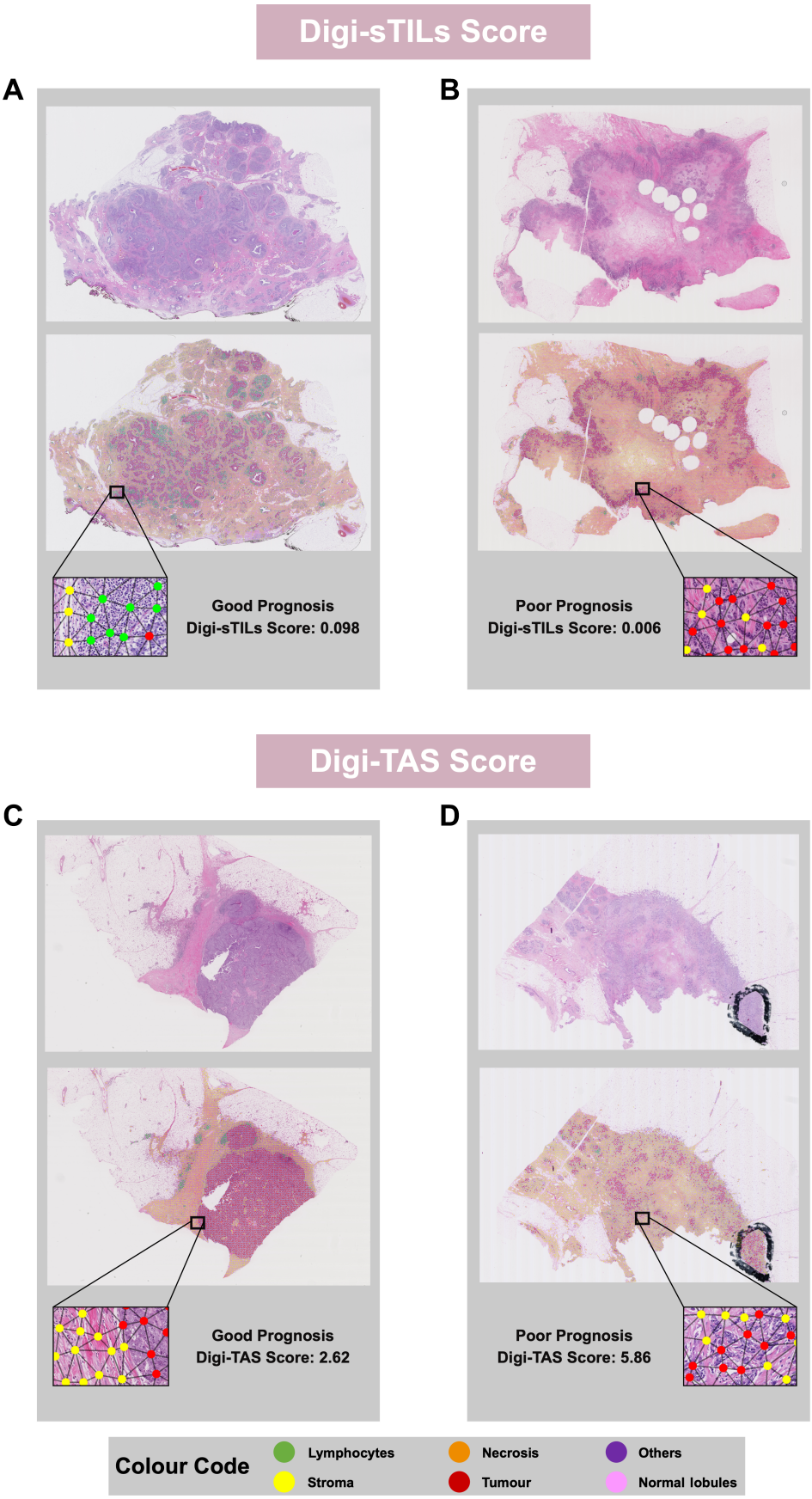

Figure 4. Illustrating the prognosis prediction difference between good and poor outcomes on the whole-slide images (WSIs).

0.71 ($p < 0.0001$ and HR of 4.0). For internal validation on the AUBC cohort and mixed cohorts, the performance of two-folds and ten runs was averaged for C-indices and HRs, and we reported twice the medians of p values. In that case, the C-indices were 0.73 and 0.68, with HR values of 6.9 ($p < 0.0001$) and 4.7 ($p < 0.0001$), respectively, for AUBC internal cross-validation and mixed cohorts, as shown in Table 3. Figure 2C shows the performance of the SurvivalSVM model using the KM curves showing clear risk stratification for time-to-event in both AUBC and TCGA cohorts using the SurvivalSVM risk prediction scores.

Discussion

There is increasing interest in the use of automated analysis of clinical and pathological data for predicting and improving patient outcomes. Traditional histopathological examinations using clinical parameters [41,42] and tumour characteristics are limited in their scope to identify and categorise specific features in the spatial tumour microenvironment and are also susceptible to intra- and inter-observer variation [43]. This study presents a novel way of automatically quantifying different features related to tissue adjacency and cellular networks using deep learning-based tissue classification. We compiled quantitative tissue descriptors concerning the spatial relationship of tumour with stroma and lymphocytes from digitised images of routine H&E sections of TNBC. When cross-validated in two independent held-out datasets, our features successfully predicted risk independent of routine clinical variables with machine learning survival models.

In our analysis, features related to sTILs and TAS were found to be strongly associated with patient outcomes in the two cohorts (AUBC and TCGA), with the strongest prognostic significance for the combination of our digital score of sTILs and the patient's positivity status for axillary lymph nodes using a machine learning-based survival model. A common theme of the proposed features is that they quantify the relationship between the tumour and its neighbouring stroma tissue and capture the arrangement of lymphocytic nodes that have direct connections with stromal nodes.

The proposed digital stromal tumour-infiltrating lymphocytes (Digi-sTILs) score indicates the ratio of stromal area occupied by lymphocytic cells to the total stroma. The C-indices of the Digi-sTILs score digital marker for disease-specific survival were between 0.60 ($p = 0.0189$) and 0.65 ($p = 0.0037$). Figure 4A shows a WSI with a good prognosis and a relatively high Digi-sTILs score, while Figure 4B shows the WSI of a case with a poor prognosis with the Digi-sTILs score being relatively low. Also, no significant departures from linearity were observed in the relationship between Digi-sTILs and DSS events, even with the addition of ALN status and age features (supplementary material, Figure S5).

Stroma tumour-infiltrating lymphocytes (sTILs), defined as lymphocytes located in the stroma but not in contact with tumour, have emerged as a robust and predictive biomarker in triple-negative breast cancer [15,44], with a relatively high value of sTILs shown to be associated with improvement in overall survival in inflammatory breast cancers. This study reinforces the role of sTILs in the context of the tumour microenvironment and eventually for the clinical outcomes of breast cancer.

Our results indicate that high-risk tumours have relatively higher levels of inter-mixing of stromal and tumour tissue as opposed to separate areas of pure stroma and tumour. The C-indices of TAS-related markers for disease-specific survival were between 0.57 ($p = 0.0437$) and 0.60 ($p < 0.0001$). Figure 4C shows a WSI with good prognosis and a relatively low Digi-TAS score, while Figure 4D shows the WSI of a case with poor prognosis with the Digi-TAS score being relatively high.

The relationship between stroma and tumour in TNBC is significant as high levels of intermixing of stroma and tumour cells are related to relatively poor prognosis compared with patients with a low tumour–stroma ratio [11,18,45], rendering our Digi-TAS score as a potential quantitative and reproducible marker for improved risk scoring. Moreover, a great deal of evidence has recently emerged showing that adjacent stromal tissue contributes significantly to regulating tumour progression. These findings present an opportunity to improve the prognostication of TNBC patients [19,46,47].

We further expanded our study with a machine learning-based model (SurvivalSVM), exploring a combination of the above features, including clinical variables, to test the utility of strengthening the model's risk stratification ability. Combining these features using the machine learning-based survival model resulted in C-indices reaching 0.76 and 0.71 in both TCGA and AUBC as validation sets, making the predictive ability of these models stronger than that of individual features. Notably, some clinical inputs were excluded due to high levels of missing data.

In summary, we have presented a novel computational approach using standard H&E histology images of TNBC and developed a set of novel digital features related to stromal tumour-infiltrating lymphocytes (sTILs) and tumour–stroma spatial arrangement (TAS) to predict survival outcomes. This study, to the best of our knowledge, is the first to show the importance of stromal features in TNBC survival outcomes with automatic quantification based on a deep learning approach. Our image-based markers successfully stratified patients into high-risk and low-risk groups. Moreover, they emphasised the role of sTILs and TAS in risk stratification. These findings could help prognosticate TNBC patients, potentially leading to new and more personalised treatment strategies with randomised control studies between our digital scores and pathologists to achieve the required high level of confidence for clinical deployment. We also investigated the prognostic value of Digi-sTILs and ALN status in addition to stage

II and compared our findings with those of Loi *et al* [16]; our findings were found to be aligned with those of Loi *et al* (supplementary material, Figure S4 and Table S8).

The proposed scores for sTILs and TAS are objective, reproducible, and prognostically significant in terms of disease-specific survival for TNBC and could assist pathologists in assessing the role of sTILs and TAS. Additional analyses using larger patient cohorts and multi-centric external validation sets are needed to confirm our findings. Integrating patient therapy types and histopathological characteristics would potentially strengthen the study findings and broaden the knowledge about survival data. The TNBC tumour subtyping is critical to better understanding the disease, and incorporating the subtype may add additional perspective on the role of tumour microenvironment biomarkers.

Author contributions statement

NR and JA conceptualised the idea. NR, JA and DG designed the study. RA wrote the code, developed the model, ran experiments and wrote the initial draft of the manuscript. JA, NP and DG provided clinical input to the study. DG arranged the data collection and JA provided manual annotations of whole-slide images. NR and FM supervised the development of the methodology. SR and MS provided support in some of the technical aspects of methodology and experiments. All authors read, modified and approved the final manuscript.

Data availability statement

Restrictions apply to the availability of the AUBC cohort, which was used under licence for this study. Data are available from the authors upon reasonable request with the permission of all parties. The TCGA cohort analysed in this study was obtained from the TCGA portal at https://tcga-data.nci.nih.gov/tcga. We have made the code for reproducing results available at the following URL: https://github.com/Rawanrb/Digi-WSIs-Analysis.git.

References

1. Balkenhol MCA, Vreuls W, Wauters CAP, *et al*. Histological subtypes in triple negative breast cancer are associated with specific information on survival. *Ann Diagn Pathol* 2020; **46:** 151490.
2. Urru SAM, Gallus S, Bosetti C, *et al*. Clinical and pathological factors influencing survival in a large cohort of triple-negative breast cancer patients. *BMC Cancer* 2018; **18:** 56.
3. Irvin WJ, Carey LA. What is triple-negative breast cancer? *Eur J Cancer* 2008; **44:** 2799–2805.
4. Gluz O, Liedtke C, Gottschalk N, *et al*. Triple-negative breast cancer – current status and future directions. *Ann Oncol* 2009; **20:** 1913–1927.
5. Dent R, Trudeau M, Pritchard KI, *et al*. Triple-negative breast cancer: clinical features and patterns of recurrence. *Clin Cancer Res* 2007; **13:** 4429–4434.
6. Vikas P, Borcherding N, Zhang W. The clinical promise of immunotherapy in triple-negative breast cancer. *Cancer Manag Res* 2018; **10:** 6823–6833.
7. Zagami P, Carey LA. Triple negative breast cancer: pitfalls and progress. *NPJ Breast Cancer* 2022; **8:** 95.
8. Yin L, Duan J-J, Bian X-W, *et al*. Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Res* 2020; **22:** 61.
9. Dieci MV, Miglietta F, Guarneri V. Immune infiltrates in breast cancer: recent updates and clinical implications. *Cell* 2021; **10:** 223.
10. Bianchini G, Balko JM, Mayer IA, *et al*. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nat Rev Clin Oncol* 2016; **13:** 674–690.
11. Vangangelt KM, Green AR, Heemskerk IM, *et al*. The prognostic value of the tumor–stroma ratio is most discriminative in patients with grade III or triple-negative breast cancer. *Int J Cancer* 2020; **146:** 2296–2304.
12. García-Teijido P, Cabal ML, Fernández IP, *et al*. Tumor-infiltrating lymphocytes in triple negative breast cancer: the future of immune targeting. *Clin Med Insights Oncol* 2016; **10:** 31–39.
13. Blackley EF, Loi S. Targeting immune pathways in breast cancer: review of the prognostic utility of TILs in early stage triple negative breast cancer (TNBC). *Breast* 2019; **48:** S44–S48.
14. Pruneri G, Vingiani A, Bagnardi V, *et al*. Clinical validity of tumor-infiltrating lymphocytes analysis in patients with triple-negative breast cancer. *Ann Oncol* 2016; **27:** 249–256.
15. Kos Z, Roblin E, Kim RS, *et al*. Pitfalls in assessing stromal tumor infiltrating lymphocytes (sTILs) in breast cancer. *NPJ Breast Cancer* 2020; **6:** 17.
16. Loi S, Drubay D, Adams S, *et al*. Tumor-infiltrating lymphocytes and prognosis: a pooled individual patient analysis of early-stage triple-negative breast cancers. *J Clin Oncol* 2019; **37:** 559–569.
17. Salgado R, Denkert C, Demaria S, *et al*. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Ann Oncol* 2015; **26:** 259–271.
18. de Kruijf EM, van Nes JG, van de Velde CJ, *et al*. Tumor–stroma ratio in the primary tumor is a prognostic factor in early breast cancer patients, especially in triple-negative carcinoma patients. *Breast Cancer Res Treat* 2011; **125:** 687–696.
19. Moorman AM, Vink R, Heijmans HJ, *et al*. The prognostic value of tumour–stroma ratio in triple-negative breast cancer. *Eur J Surg Oncol* 2012; **38:** 307–313.
20. Hagenaars SC, Vangangelt KM, Van Pelt GW, *et al*. Standardization of the tumor–stroma ratio scoring method for breast cancer research. *Breast Cancer Res Treat* 2022; **193:** 545–553.
21. Ibrahim A, Gamble P, Jaroensri R, *et al*. Artificial intelligence in digital breast pathology: techniques and applications. *Breast* 2020; **49:** 267–273.
22. Deng L, Yu D. Deep Learning: Methods and Applications. In *Foundations and Trends® in Signal Processing* (Vol. **7**). now Publishers Inc.: Hanover, MA, 2014; 197–387.
23. Litjens G, Kooi T, Bejnordi BE, *et al*. A survey on deep learning in medical image analysis. *Med Image Anal* 2017; **42:** 60–88.
24. Graham S, Vu QD, Raza SEA, *et al*. Hover-net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med Image Anal* 2019; **58:** 101563.
25. Sirinukunwattana K, Ahmed Raza SE, Tsang Y-W, *et al*. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Trans Med Imaging* 2016; **35:** 1196–1206.
26. Awan R, Koohbanani NA, Shaban M, *et al*. Context-aware learning using transferable features for classification of breast cancer histology images. In *International Conference Image Analysis and Recognition* (Vol. **1**). Springer: Cham, 2018; 788–795.
27. Ker J, Bai Y, Lee HY, *et al*. Automated brain histology classification using machine learning. *J Clin Neurosci* 2019; **66:** 239–245.

28. Balkenhol MC, Bult P, Tellez D, *et al*. Deep learning and manual assessment show that the absolute mitotic count does not contain prognostic information in triple negative breast cancer. *Cell Oncol* 2019; **42:** 555–569.
29. 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.
30. Saltz J, Gupta R, Hou L, *et al*. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep* 2018; **23:** 181–193.e7.
31. Albusayli R, Graham D, Pathmanathan N, *et al*. Simple non-iterative clustering and CNNs for coarse segmentation of breast cancer whole-slide images. In *Medical Imaging 2021: Digital Pathology* (Vol. **11603**). SPIE, San Diego, CA, 2021; 100–108.
32. He K, Zhang X, Ren S, *et al*. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE: Las Vegas, NV, 2016; 770–778.
33. Litjens G. Automated Slide Analysis Platform. [Accessed 19 October 2019]. Available from: https://github.com/computationalpathologygroup/ASAP.
34. Achanta R, Susstrunk S. Superpixels and polygons using simple non-iterative clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE: Honolulu, HI, 2017; 4651–4660.
35. Steck H, Krishnapuram B, Dehing-Oberije C, *et al*. On ranking in survival analysis: bounds on the concordance index. In *Proceedings of the 20th International Conference on Neural Information Processing Systems*. Curran Associates Inc.: Vancouver, 2007; 1209–1216.
36. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996; **15:** 361–387.
37. Uno H, Cai T, Pencina MJ, *et al*. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med* 2011; **30:** 1105–1117.
38. Schröder MS, Culhane AC, Quackenbush J, *et al*. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics* 2011; **27:** 3206–3208.
39. Pölsterl S, Navab N, Katouzian A. Fast training of support vector machines for survival analysis. In *Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2015. Lecture Notes in Computer Science, Appice A, Rodrigues P, Santos Costa V,* et al *(eds), Vol 9285*. Springer: Cham, 2015; 243–259.
40. Romano JP, DiCiccio C. *Multiple data splitting for testing*. Technical Report No. 2019-03. Department of Statistics, Stanford University: Stanford, CA, 2019 Available from: https://purl.stanford.edu/fb041jg0790.
41. Rakha EA, El-Sayed ME, Green AR, *et al*. Prognostic markers in triple-negative breast cancer. *Cancer* 2007; **109:** 25–32.
42. Liedtke C, Hess KR, Karn T, *et al*. The prognostic impact of age in patients with triple-negative breast cancer. *Breast Cancer Res Treat* 2013; **138:** 591–599.
43. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. *Lancet Oncol* 2019; **20:** e253–e261.
44. Van Berckelaer C, Rypens C, van Dam P, *et al*. Infiltrating stromal immune cells in inflammatory breast cancer are associated with an improved outcome and increased PD-L1 expression. *Breast Cancer Res* 2019; **21:** 28.
45. Kramer CJH, Vangangelt KMH, van Pelt GW, *et al*. The prognostic value of tumour–stroma ratio in primary breast cancer with special attention to triple-negative tumours: a review. *Breast Cancer Res Treat* 2019; **173:** 55–64.
46. Beck AH, Sangoi AR, Leung S, *et al*. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci Transl Med* 2011; **3:** 108ra113.
47. Conklin MW, Keely PJ. Why the stroma matters in breast cancer: insights into breast cancer patient outcomes through the examination of stromal biomarkers. *Cell Adh Migr* 2012; **6:** 249–260.

SUPPLEMENTARY MATERIAL ONLINE

Figure S1. A chart that shows the criteria used to include and exclude whole-slide images in our study from both cohorts (AUBC, TCGA)

Figure S2. The Kaplan–Meier estimates of the two survival endpoints at 10 years according to the subgroups defined by grade (grade II and grade III)

Figure S3. The Digi-sTILs scores in the TCGA cohort were positively correlated using Spearman's correlation coefficient with three pathologists' manual stroma TILs scoring

Figure S4. Forest plots showing the results of multivariate regression analysis for each feature and the clinical inputs in the AUBC cohort, with 318 cases for A and 311 cases for B and C as the ungraded and with no tumour size information cases, are excluded

Figure S5. Relationship between Digi-sTILs treated as a continuous variable and modelled using restricted cubic splines, and the relative risk of DSS event with adding of ALN status and age features

Table S1. TCGA-TNBC slide identities and their Digi-sTILs and Digi-TAS scores, which the study used for cross-validation testing

Table S2. Distribution of annotated patches for each class in both training and validation sets for our segmentation model training and validation phases

Table S3. Performance of our segmentation model for each tissue type

Table S4. Description and mathematical equations of our digital features

Table S5. Univariate disease-specific survival analysis and survival machine learning model were performed on the two cohorts using the studied features

Table S6. Univariate disease-specific survival analysis with the survival machine learning models was performed on the AUBC (two-folds, ten runs)

Table S7. Univariate disease-specific survival analysis with the survival machine learning models was performed on the AUBC–TCGA mixed cohorts (two-folds, ten runs)

Table S8. Kaplan–Meier estimates of the two survival endpoints at 10 years according to the subgroups defined by Digi-sTILs, with a cut-off defined based on the dataset used for training, axillary lymph node positivity status, and stage II