



Research paper

A Computational Tumor-Infiltrating Lymphocyte Assessment Method Comparable with Visual Reporting Guidelines for Triple-Negative Breast Cancer



Peng Sun^{a,b,#,*}, Jiehua He^{a,b,#}, Xue Chao^{a,b,#}, Keming Chen^{a,b}, Yuanyuan Xu^c, Qitao Huang^{a,b}, Jingping Yun^{a,b}, Mei Li^{a,b}, Rongzhen Luo^{a,b}, Jinbo Kuang^d, Huajia Wang^d, Haosen Li^d, Hui Hui^d, Shuoyu Xu^{d,e,*,**}

^a State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangzhou, P. R. China

^b Department of Pathology, Sun Yat-sen University Cancer Center, Guangzhou, P. R. China

^c Department of Physiology, Zhongshan School of Medicine, Sun Yat-Sen University, Guangzhou, P. R. China

^d Bio-totem Pte Ltd, Foshan, P. R. China

^e Department of General Surgery, Nanfang Hospital, Southern Medical University, Guangzhou, P. R. China

ARTICLE INFO

Article History:

Received 31 March 2021

Revised 30 June 2021

Accepted 30 June 2021

Available online xxx

Keywords:

Triple-negative breast cancer
Tumor-infiltrating lymphocyte
Deep learning
Prognosis

ABSTRACT

Background: Tumor-infiltrating lymphocytes (TILs) are clinically significant in triple-negative breast cancer (TNBC). Although a standardized methodology for visual TILs assessment (VTA) exists, it has several inherent limitations. We established a deep learning-based computational TIL assessment (CTA) method broadly following VTA guideline and compared it with VTA for TNBC to determine the prognostic value of the CTA and a reasonable CTA workflow for clinical practice.

Methods: We trained three deep neural networks for nuclei segmentation, nuclei classification and necrosis classification to establish a CTA workflow. The automatic TIL (aTIL) score generated was compared with manual TIL (mTIL) scores provided by three pathologists in an Asian ($n = 184$) and a Caucasian ($n = 117$) TNBC cohort to evaluate scoring concordance and prognostic value.

Findings: The intraclass correlations (ICCs) between aTILs and mTILs varied from 0.40 to 0.70 in two cohorts. Multivariate Cox proportional hazards analysis revealed that the aTIL score was associated with disease free survival (DFS) in both cohorts, as either a continuous [hazard ratio (HR)=0.96, 95% CI 0.94–0.99] or dichotomous variable (HR=0.29, 95% CI 0.12–0.72). A higher C-index was observed in a composite mTIL/aTIL three-tier stratification model than in the dichotomous model, using either mTILs or aTILs alone.

Interpretation: The current study provides a useful tool for stromal TIL assessment and prognosis evaluation for patients with TNBC. A workflow integrating both VTA and CTA may aid pathologists in performing risk management and decision-making tasks.

© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

1. Introduction

Recent evidence has suggested that tumor-infiltrating lymphocytes (TILs) have prognostic and predictive capabilities for triple-

negative (TNBC) and human epidermal growth factor receptor 2 (HER2)-positive breast cancers [1–7]. TIL assessments are clinically useful for risk predictions, adjuvant and neoadjuvant chemotherapy decisions, and more recently, immunotherapy [8]. TIL assessments are being included in clinical trials and diagnostic assessments, which has raised concerns regarding the existence of a standardized methodology for evaluating TILs. Therefore, recommendations and guidelines for visual TIL assessment (VTA) in invasive breast carcinoma patients have been recently developed by the International Immunology Biomarker Working Group (or TILs-WG) [9,10]. TIL populations are semiquantified by determining how much of a demarcated area of stroma or tumor visible on a slide is infiltrated by immune cells (average TIL%). Although several studies with small sample sizes

* Corresponding authors: Dr. Peng Sun, Department of Pathology, Sun Yat-sen University Cancer Center, 651 Dongfeng East Road, 510060, Guangzhou, P. R. China

** Co-corresponding author: Dr. Shuoyu Xu: Bio-totem Pte Ltd, 321 State Road, 528231, Foshan, P.R.China; Department of General Surgery Nanfang Hospital, Southern Medical University, 1838 Guangzhou Avenue North, 510515, Guangzhou, P.R.China.

*** 1838 Guangzhou Avenue North, 510515, Guangzhou, P. R. China

E-mail addresses: sunpeng1@sysucc.org.cn (P. Sun), shuoyu.xu@bio-totem.com (S. Xu).

Peng Sun, Jiehua He and Xue Chao contributed equally to this paper as co-first authors.

Research in context

Evidence before this study

Tumor-infiltrating lymphocytes (TILs) are important prognostic biomarkers in triple negative breast cancer (TNBC) as well as in many other types of solid tumors. The International Immunology Biomarker Working Group (TIL-WG) has established a detailed reporting guideline for visual assessment of TILs (VTA) for TNBC which helps to reduce manual scoring variations but still exists inherent limitations. The development of new computational TIL assessment (CTA) methods is a promising solution to address these limitations. We searched articles from Pubmed database with the terms ("tumor infiltrating lymphocytes") AND ("TNBC" OR "breast cancer") AND ("deep learning" OR "artificial intelligence" OR "computational assessment"). The found CTA methods established alternative quantitative metrics (eg. lymphocyte percentage, spatial patterns of lymphocyte distribution) for stromal TILs assessment rather than being consistent with VTA guideline. Moreover, all these methods were validated in Caucasian cohort only.

Added value of this study

The CTA method in this study is developed broadly following the visual reporting guidelines to quantify area percentage of TILs-dense areas over all the tumor stromal areas, which enables a direct comparison with VTA from multiple pathologists. The automated TILs (aTILs) score was found to be consistent with manual TILs (mTILs) score, and was associated with disease-free survival in both Asian and Caucasian cohorts as either a continuous or dichotomous variable. We discovered that the optimized TILs cut-off values for patient stratification varied between different ethnicities. Our results also suggested a composite model by combining both aTILs and mTILs could help junior pathologist for better risk stratification.

Implications of all the available evidence

The proposed CTA method provides a useful stromal TIL scores and prognostic stratification in TNBCs, which is consistent with VTA. Besides, a workflow integrating both VTA and CTA is illustrated which may aid pathologists in performing risk management and decision-making tasks in clinical practice.

[11–15] have been conducted since then, demonstrating acceptable agreement among observers for decentralized TIL assessments of breast cancer, there are still difficulties in performing consistent evaluations on the basis of hematoxylin and eosin (H&E)-stained sections due to factors such as individual outliers or cutoff points, the spatial distribution of TILs, the tumor-stroma ratio, histologic subtypes, TILs in ductal carcinomas in situ, the perivascular abundance of TILs, necrosis, the tertiary lymphoid structure (TLS), and intratumoral heterogeneity. All of these factors may increase the interobserver and intraobserver variability of VTAs [8].

Ring studies by Denkert et al. [11] have led to the development of a preliminary software-guided TIL evaluation that addresses variation caused by intratumoral heterogeneity and random errors, which resulted in an improved concordance rate among pathologists. In addition, an increasing number of studies have used machine learning algorithms to analyze histological images and have recently shown promising results regarding diagnostics, therapeutic target predictions, and prognostic evaluations [9,16,17]. These successes may suggest that quantitative, automated, and reproducible computational TIL assessments (CTAs) [18] of whole-slide images (WSIs) based on machine learning algorithms may be useful for further

standardization. For breast cancer, various automated computerized tools have been previously developed to detect lymph node metastases [19], identify tumor-associated stroma [20], classify mitoses [21], or predict molecular subtypes [22]. A few studies have been conducted to explore TILs based on histologic H&E images using deep learning models and breast cancer data. Saltz et al. [23] reported the spatial organization and molecular correlations of TILs using deep learning and pathology images of 13 tumor types from The Cancer Genome Atlas (TCGA) database. Moreover, Abe et al. [24] preliminarily performed a quantitative digital image analysis of TIL density, the number of TILs per unit area, in HER2-positive breast cancer cases. These previous studies demonstrate the potential to improve the precision of histomorphological evaluations of TILs, thereby promoting its implementation in clinical decision-making.

In the present study, we developed a CTA algorithm for automatic stromal TIL quantification (aTILs) in TNBC cases based on WSIs. We also retrospectively reviewed the data of patients diagnosed with TNBC at our cancer center (Sun Yat-sen University Cancer Center (SYSUCC), $N = 184$) and from the TCGA database ($N = 117$), which were included as validation cohort data for the CTA model. Clinicopathologic features, such as age, tumor size, lymph node metastasis (LNM), the TNM stage, clinical treatment, and survival outcomes, were evaluated. The coefficient of variation (CV), intraclass correlation coefficient (ICC), and Cohen's kappa score were assessed to estimate the interobserver agreement among three pathologists in the manually determined stromal TILs (mTILs), as well as the concordance between mTIL and aTIL scores. We subsequently compared mTILs and aTILs in terms of prognostic value in both cohorts using the Kaplan–Meier survival curve and Cox proportional hazards regression model for disease-free survival (DFS). In addition, we developed a composite mTIL/aTIL three-tier stratification model and compared its prognostic performance with that of a two-tier stratification model using mTILs or aTILs. The objectives of the present study were to generate a CTA algorithm that can precisely and automatically calculate stromal TILs in TNBC cases and to explore its prognostic value and potential application in clinical practice.

2. Materials and methods

2.1. Identification of the SYSUCC cohort

Clinical data, including data on the age at diagnosis, tumor size, LNM, TNM stage, histological grade, clinical treatments, and survival outcomes, for patients with pathologically confirmed invasive breast cancer in the SYSUCC during 2005–2010 were collected. All patients underwent operations at SYSUCC, and formalin-fixed paraffin-embedded (FFPE) tissue specimens, including the tumor, sentinel lymph nodes, and axillary lymph nodes, were stained routinely with H&E. Tumor staging was performed on the basis of the criteria established by the 8th edition of the American Joint Committee on Cancer (AJCC 8th) staging manual. Estrogen receptor (ER), progesterone receptor (PR), and HER2 statuses were determined by immunohistochemical (IHC) staining. ER and PR statuses were classified as negative if the values were below the cutoff of 1%, according to the American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP) guidelines [25]. HER2 status was defined as negative, with 0, 1+ as well as 2+ on IHC without HER2 gene amplification on fluorescence in situ hybridization (FISH) [26]. The archived H&E and IHC slides were retrospectively reviewed by three pathologists (HJH, SP, and CX) to confirm the diagnosis. A total of 184 patients with TNBC were included in the SYSUCC cohort.

2.2. Identification of TCGA cohort

The TCGA cohort was identified from the public Cancer Genome Atlas Breast Invasive Carcinoma (TCGA_BRCA) database. The clinical

data, including age at diagnosis, TNM stages, histological grades, and survival outcomes, of 1097 patients with invasive breast cancer were retrieved from the TCGA Pan-Cancer Clinical Data Resource [27]. ER, PR and HER2 status was collected from TCGA data portal. ER-, PR-, or HER2-positive/unknown patients and those without H&E-stained WSIs available from FFPE samples were excluded, and a total of 117 patients with TNBC were included in the TCGA cohort.

2.3. Sample preparation and whole-slide image acquisition

One representative FFPE tissue block of breast lesions was collected for individuals with TNBC in the SYSUCC cohort. The block was sectioned into 4- μm -thick tissue slides for H&E staining (VENTANA HE 600 system, Roche). All slides were scanned at 40x magnification, with an image resolution of 0.25 $\mu\text{m}/\text{pixel}$ (Aperio, ScanScope AT2, Leica). The scanned WSIs from the TCGA cohort were directly downloaded from the NIH portal (portal.gdc.cancer.gov), with a resolution of either 0.25 $\mu\text{m}/\text{pixel}$ or 0.5 $\mu\text{m}/\text{pixel}$. We also collected an additional 148 tissue H&E slides (training set) from TNBC patients diagnosed at the SYSUCC, who were not included in the SYSUCC cohort, for image analysis algorithm training. These slides were prepared, stained, and scanned using the same equipment and settings mentioned above.

2.4. Manual stromal TIL evaluation and annotation

All WSIs from the SYSUCC and TCGA cohorts were evaluated independently by three experienced breast pathologists. The pathologists had a diagnostic experience of 15 years (mTILs-1, HJH), 8 years (mTILs-2, SP), or 3 years (mTILs-3, CX) using ImageScope software (Leica Biosystems) to score mTILs according to a five-step standardized scoring system developed by the International Immunology Biomarker Working Group [9,10]. In brief, TILs encompass all mononuclear cells (including lymphocytes and plasma cells), and polymorphonuclear leukocytes (neutrophils) were excluded. The percentage of stromal TILs in the entire studied area (stromal TILs%) was assessed within the invasive tumor's borders. TILs in tumor areas with crush artifacts, necrosis, and regressive hyalinization and those in previous biopsy sites were excluded. The percentage of stromal TILs was considered a semiquantitative continuous parameter indicating how much of the demarcated stromal area exhibits dense mononuclear infiltrates. One pathologist (SP) marked the contours of necrosis regions on each WSI from the training cohort and annotated the different cell types (malignant epithelial cells, mononuclear cells, and other cells) in 740 cropped images measuring 1024 \times 1024 pixels. These annotations were then used for training and validation of the image analysis algorithms.

2.5. Whole-slide image analysis

Regarding the quantification of TILs in the stromal regions, the key is to identify both stromal areas and mononuclear cells from WSIs. In previous studies [23], this step was achieved by the direct classification of image patches as tumor, stroma, or lymphocyte regions. However, one patch might contain different tissue components, which makes classification difficult. Moreover, this patch-based approach may not provide detailed, quantitative information on the number or density of TILs. Alternatively, we used a cell-based approach in this study; we identified the nucleus of each cell type, including malignant epithelial cells and mononuclear cells, in WSIs (Fig. 1a,b). The stromal areas could also be accurately recognized from nuclei density maps. Because deep learning models have been shown to have significant advantages in a large number of computer vision tasks, we used several deep learning networks for nuclei segmentation, nuclei classification, and necrosis detection.

2.6. Nuclei segmentation

To first localize the nuclei in WSIs, we used a nuclei segmentation model that was developed with the MICCAI multiorgan nuclei segmentation (MoNuSeg) challenge dataset [28], which contains 30 well-annotated images extracted from H&E slides of various types of organs. Our approach mainly relied on the Mask-RCNN network [29], which is one of the most popular segmentation architectures used in the challenge. We selected a 101-layer deep ResNet [30] network as the convolutional backbone in the Mask-RCNN network. We modified the regional proposal network anchor area scale to better fit the nuclei segmentation task because the nuclei in the WSIs were small. The training parameters were carefully tuned, and the test time was increased when the segmentation masks were generated. The aggregated Jaccard index (AJI) of our segmentation model was 0.665, 0.610, and 0.662 on the training set, validation set, and unseen test set images, respectively, and the model ranked 9th in the challenge. Methods and training details can also be found in our previous study [31].

2.7. Nuclei classification

To create a training set for nuclei classification, we extracted 740 images with a size of 1024 \times 1024 pixels from the 148 WSIs in the training cohort. One pathologist (SP) carefully annotated the cell types. A total of 61,620 malignant epithelial cells, 38,154 mononuclear cells, and 33,729 other cells were annotated and split into the training, validation, and testing sets at a proportion of 8:1:1 while the ratio between three types of cells was same in each set.

Several CNN models, such as ResNet, Inception, Xception, and NASNetLarge, have shown excellent performance in various classification tasks. In this study, we selected the Xception architecture [32], with initial weights that were pretrained on ImageNet for nuclei classification, as this network has been demonstrated to outperform other state-of-art networks in classifying histological breast cancer images in a recent study [33]. Extensive data augmentation, such as rotating, flipping, cropping, and strong color augmentation in HSV color space, was performed during training. The model was trained for at least 20 epochs using the stochastic gradient descent (SGD) optimizer and categorical cross-entropy loss function with a learning rate of 10^{-4} .

The size of a nucleus varies significantly, and the surrounding areas of the nuclei might be useful in improving the classification results. Thus, we adopted a multiresolution approach and extracted nucleus patches of four different sizes (27 \times 27, 36 \times 36, 45 \times 45, and 54 \times 54 pixels at 40 \times amplification) from each nucleus center and trained an individual Xception network according to each patch size. The prediction of four networks was integrated by averaging the prediction probability of each cell class. The average weighted F1 score and Matthews Correlation Coefficient (MCC) were 0.856 and 0.770 for the validation set and 0.851 and 0.765 for testing set, respectively, showing excellent performance for nuclei classification (**Supplementary Table 1**).

2.8. Necrosis detection

All necrotic areas were annotated in WSIs from the training cohort, and the 224 \times 224 pixel patches from the necrotic regions were extracted at the highest magnification of 40 \times , which accounted for a total of 139,909 patches. Additionally, 198,274 patches of the same size were randomly sampled from the nonnecrotic regions. All patches were split into training, validation, and testing sets at a ratio of 8:1:1, and the ratio between necrosis and non-necrosis patches was kept same at around 0.7:1 for three sets. As with nuclei classification, we adopted the Xception network for the classification of necrosis and nonnecrotic patches. The data augmentation and training

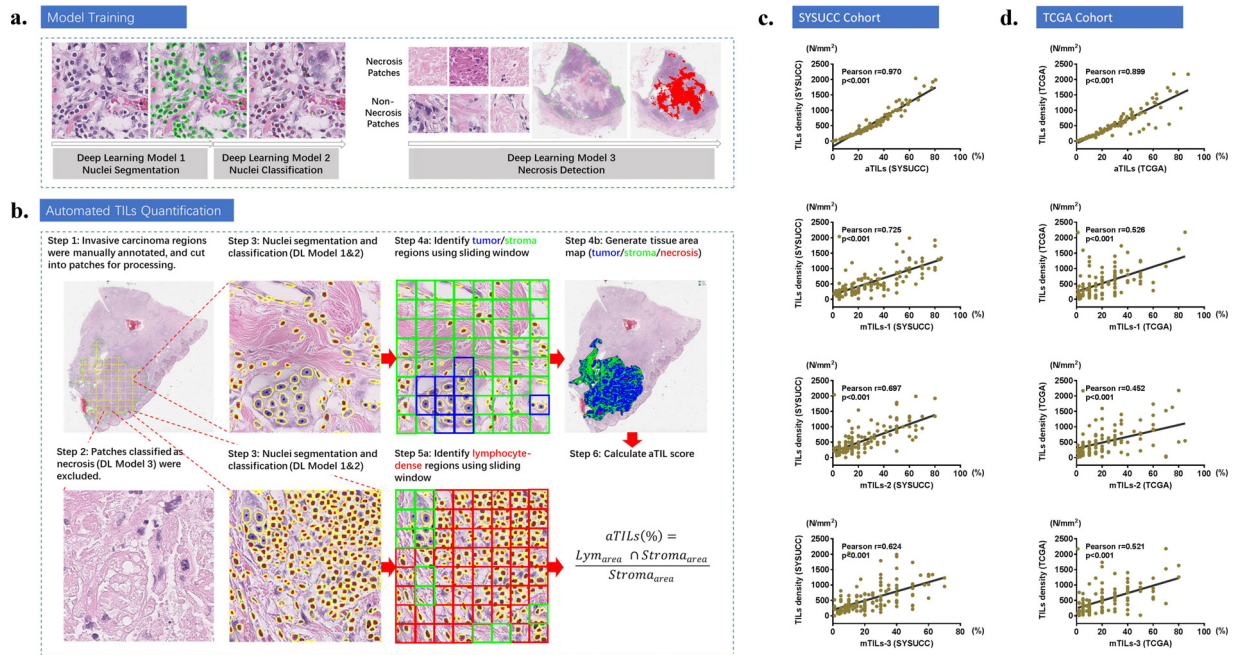


Fig. 1. Overview of the deep learning model used for CTA and the correlation between TIL density and TIL scores. (a) Three deep learning models were trained for nuclei segmentation, nuclei classification, and necrosis detection. (b) A flowchart of the automated stromal TIL quantification process. Step 1: The invasive tumor area was manually outlined by the pathologist (green contour). Step 2: The regions inside the outlined area were cut into 224×224 pixels patches and classified into necrotic and non-necrotic; the necrotic patches were excluded for the following analysis. Step 3: The nuclei in each non-necrosis patch were segmented (yellow contours) and classified into tumor (blue dots), lymphocyte (red dots) and others (no dot). Step 4: A sliding window of 128×128 pixels corresponding to a size of $32 \times 32 \mu\text{m}$ was used to visualize all of the non-necrotic regions; if the number of tumor cell nuclei in the window was greater than 2, then this region was considered a tumor region (blue blocks). The remaining regions were recognized as the stromal region (green blocks). The identified tumor, stroma and necrosis regions in the entire WSI were also shown. Step 5: A smaller 32×32 pixel ($8 \times 8 \mu\text{m}$) sliding window was used to view all of the stromal regions; if the number of lymphocytes in the window was greater than 2, then the region was considered lymphocyte-dense (red blocks). Step 6: The final aTIL score was calculated as the overlapping area between lymphocyte-dense regions and stromal regions divided by the area of the stromal regions. The sliding window without overlap was shown for illustration purpose, as 50% overlapping was actually adopted. (c, d) The stromal TIL density within the invasive tumor area positively correlated with the mTIL and aTIL scores in both cohorts.

parameters were also similar to those used in the nuclei classification task. The trained network yielded weighted F1 scores and MCC of 0.900 and 0.798 for the validation set and 0.818 and 0.801 for the testing set, respectively (**Supplementary Table 1**).

The visual assessment of the outputs of three models from all the images in SYSUCC and TCGA cohort were performed by a pathologist (SP) to ensure no obvious segmentation and classification mistakes were found from a large number of nuclei or areas thus the results were suitable for downstream analysis.

2.9. Stromal til quantification

One straightforward method of quantifying stromal TILs is to calculate the density of mononuclear cells in stromal areas. The overall invasive breast carcinoma area was first annotated by a pathologist (SP) in the WSIs to exclude the benign and ductal carcinoma in situ (DCIS) regions. The necrotic areas within the invasive carcinoma area were then detected using the trained deep learning network. All of the nuclei in the nonnecrotic invasive carcinoma area were segmented and then classified to generate nuclei density maps of malignant epithelial cells and mononuclear cells using a sliding window approach. Specifically, a sliding window of 128×128 pixels moved in steps of 64 pixels among nonnecrotic invasive carcinoma areas and counted the number of malignant epithelial cells in each window. Steps of 64 pixels were used to increase the following tumor/stroma region image resolution and to avoid missing nuclei at the boundary of each sliding window. If the number of malignant epithelial cells in the current window was greater than the set threshold, the block was identified as a tumor region; otherwise, it was considered a stromal region. The threshold was empirically set to 2 in this study. The density of mononuclear cells in stromal areas could then

be quantified as the total number of TILs in the stromal regions divided by the entire stromal region.

However, the above density-based measurement was different from the definition used for manual sTIL scoring, making it difficult to compare the two methods. Thus, we constructed a new computational metrics, the percentage of TIL-dense stroma blocks within all stroma blocks, using the sliding window method. A TIL-dense stroma block was defined as a block with more than two mononuclear cells which is equivalent to the stroma area occupied by TILs in our definition. This metrics ranged between 0 and 100% and highly correlated with the mononuclear cell density-based measurement (SYSUCC cohort: $r = 0.97$, 95% CI 0.96–0.98, $p < 0.001$; TCGA cohort: $r = 0.90$, 95% CI 0.86–0.93, $p < 0.001$; t -test; **Fig. 1c**). In all subsequent analyses in this study, we used the above metrics as the automated quantified stromal TIL score (aTILs). The image analysis and quantification processes are illustrated in **Fig. 1a,b**.

3. Ethics

This study has been approved by the Sun Yat-sen University Cancer Center (SYSUCC) ethics committee (reference number: 047/20). The requirement to obtain informed consent from the participants was waived by the ethics committee.

3.1. Statistical analysis

The CV was used to characterize the variation in manual scores for each sample. The Pearson correlation coefficient was used to analyze the correlation between TIL density and the aTIL score, between TIL density and the CV, as well as between the stroma ratio and CV. The difference in TIL density and the stroma ratio between groups was

compared using Student's *t*-test. Intraclass correlation (ICC) analysis was used to estimate the interobserver agreement among three pathologists in stromal TIL evaluations, as well as the differences between aTIL and mTIL scores. There are currently no formal recommendations for a clinically relevant cutoff point for stromal TILs. We used various cutoff points, including 10%, 20%, 30%, and 40%, as reported in previous studies [3–6,8], to stratify the patients into two groups (TILs-High vs. TILs-Low) and calculated the Cohen's kappa score to assess the agreement between mTIL and aTIL scores. For survival analysis, the primary endpoint used was DFS, which was defined as the first relapse or death from any cause. X-tile software (version 3.6.1; Yale University, New Haven, CT, USA) was used to identify the optimal cutoff mTIL and aTIL scores for DFS. The survival curves of DFS were drawn using the Kaplan–Meier method and were compared using log-rank tests.

Univariate and multivariate analyses of DFS were performed using the Cox proportional hazards regression model, and hazard ratios (HRs) were also calculated. The Harrell concordance index (C index) was calculated to measure the model's predictive performance for DFS using different TIL scoring methods. A higher C index indicated a better predictive performance. All tests were two-sided, and a *p*-value <0.05 was considered significant. Statistical analyses were performed using Medcalc software v19.0.4 (MedCalc Software Ltd., Belgium), MATLAB software v2017b (The MathWorks, Inc., Natick, MA, USA), and RStudio software v1.2.5 (RStudio, Inc.).

3.2. Role of funders

The funder had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

4. Results

4.1. Patient characteristics

The characteristics of the patients with TNBC in the SYSUCC and TCGA cohorts are summarized in Table 1. In the SYSUCC cohort, the median age of the patients at diagnosis of TNBC was 48 years (range, 23–80 years). The median follow-up period was 53.5 months (range, 1–271 months). Disease-related events (either relapse, metastasis, or death) occurred in 23.9% of the patients. The median age of patients at diagnosis of TNBC was 53 years (range, 26–90 years) in the TCGA cohort, with a median follow-up of 30.5 months (range, 0.2–286 months). A total of 19.7% of the patients in the TCGA cohort had disease-related events.

4.2. Interobserver agreement among mTILs

A heatmap of the aTIL and mTIL scores is shown in Fig. 2a. The CV of mTIL scores among the three pathologists was calculated to evaluate CTA consistency. As shown in Fig. 2b, a negative correlation was observed between the automatic measurements of the stromal TIL density and CV in both cohorts (SYSUCC cohort: Pearson $r = -0.16$, $p = 0.027$; TCGA cohort: Pearson $r = -0.30$, $p = 0.001$; *t*-test), while no correlations were found between the automatic measurements of the tumor-stroma ratio (TSR) and CV. The median CV values of the mTIL scores were 0.29 [interquartile range (IQR), 0.17–0.45] and 0.43 (IQR, 0.29–0.66) in the SYSUCC and TCGA cohorts, respectively (Fig. 2c). The stromal TIL density was significantly higher (SYSUCC cohort, $609.9 \pm 43.1/\text{mm}^2$ vs. 430.9 ± 39.7 , $p = 0.003$; TCGA cohort, $633.8 \pm 62.7/\text{mm}^2$ vs. $439.2 \pm 54.1/\text{mm}^2$, $p = 0.020$; student's *t*-test) in the CV-low cases than in the CV-high cases in both cohorts (Fig. 2d). In addition, no significant differences in TSR were found between CV-low and CV-high cases in either the SYSUCC cohort

Table 1
Characteristics of patients with TNBC in SYSUCC and TCGA cohorts.

Variable	SYSUCC cohort (N = 184)	TCGA cohort (N = 117)
Age at diagnosis (y)		
<40	34 (18.5)	14 (12.0)
40–55	107 (58.1)	50 (42.7)
>55	43 (23.4)	53 (45.3)
Histological grade		
Grade II	21 (11.4)	13 (11.1)
Grade III	163 (88.6)	104 (88.9)
Tumor size (pT)		
pT1	24 (13.0)	24 (20.5)
pT2	130 (70.7)	78 (66.7)
pT3	30 (16.3)	11 (9.4)
pT4	0 (0.0)	4 (3.4)
LNM (pN)		
pN0	93 (50.5)	75 (64.1)
pN1	72 (39.1)	27 (23.1)
pN2	17 (9.2)	9 (7.7)
pN3	2 (1.1)	6 (5.1)
TNM staging		
I	13 (7.1)	24 (20.5)
II	146 (79.3)	76 (65.0)
III	25 (13.6)	17 (14.5)
Chemotherapy		
No	18 (9.8)	–
Yes	166 (90.2)	–
Radiotherapy		
No	155 (84.2)	–
Yes	29 (15.8)	–
Disease related events		
Yes	44 (23.9)	23 (19.7)
No	140 (76.1)	94 (80.3)

TNBC, triple negative breast cancer; LNM, lymph node metastasis.

(45.6 ± 1.4 vs. 47.7 ± 1.8 , $p = 0.354$, student's *t*-test) or the TCGA cohort (48.4 ± 2.2 vs. 47.4 ± 2.2 , $p = 0.735$, student's *t*-test) (Fig. 2d).

The intraclass correlations (ICCs) of the mTIL scores in the SYSUCC cohort among the three pathologists were 0.89 (95% CI 0.85–0.91, mTILs-1 vs. mTILs-2), 0.81 (95% CI 0.75–0.85, mTILs-1 vs. mTILs-3), and 0.85 (95% CI 0.79–0.88, mTILs-2 vs. mTILs-3), which indicated excellent interobserver agreement (Fig. 3a). However, there was moderate agreement in the mTIL scores among the pathologists in the TCGA cohort, as the ICCs were lower: 0.61 (95% CI 0.48–0.71, mTILs-1 vs. mTILs-2), 0.66 (95% CI 0.53–0.74, mTILs-1 vs. mTILs-3), and 0.68 (95% CI 0.56–0.76, mTILs-2 vs. mTILs-3) (Fig. 3d).

4.3. Concordance between mTILs and aTILs

The ICCs between the aTILs and mTILs scores varied from 0.62 to 0.70 in the SYSUCC cohort (aTILs vs. mTILs-1, ICC=0.70, 95% CI 0.61–0.76; aTILs vs. mTILs-2, ICC=0.68, 95% CI 0.59–0.75; aTILs vs. mTILs-3, ICC=0.62, 95% CI 0.52–0.70), while they were lower, ranging from 0.40 to 0.52, in the TCGA cohort (aTILs vs. mTILs-1, ICC=0.52, 95% CI 0.37–0.64; aTILs vs. mTILs-2, ICC=0.40, 95% CI 0.23–0.53; aTILs vs. mTILs-3, ICC=0.45, 95% CI 0.28–0.57). Both cohorts showed fair-to-moderate agreement between the VTA and CTA. Among the CV-low cases, the ICCs improved slightly to 0.66–0.70 and 0.66–0.72 in the SYSUCC and TCGA cohorts, respectively (Fig. 3b and 3e). Although decreased ICCs were observed in the CV-high cases in both cohorts (0.48–0.68 in the SYSUCC cohort; 0.20–0.40 in the TCGA cohort), the mTIL scores provided by senior pathologists were shown to be more consistent with the aTIL scores (mTILs-1 vs. aTILs: ICC=0.68, 95% CI 0.55–0.77 in the SYSUCC cohort; ICC=0.37, 95% CI 0.14–0.56 in the TCGA cohort) than the mTIL scores provided by junior pathologists (mTIL-1 vs. mTIL-3: ICC=0.55, 95% CI 0.38–0.67 in the SYSUCC cohort; mTIL-1 vs. mTIL-3: ICC=0.34, 95% CI 0.11–0.54 in the TCGA cohort) for these CV-high cases (Fig. 3c and 3d). Scatter plots between mTILs and aTILs scores were also summarized in Supplementary Figure 1 with lower RMSE found in SYSUCC cohort

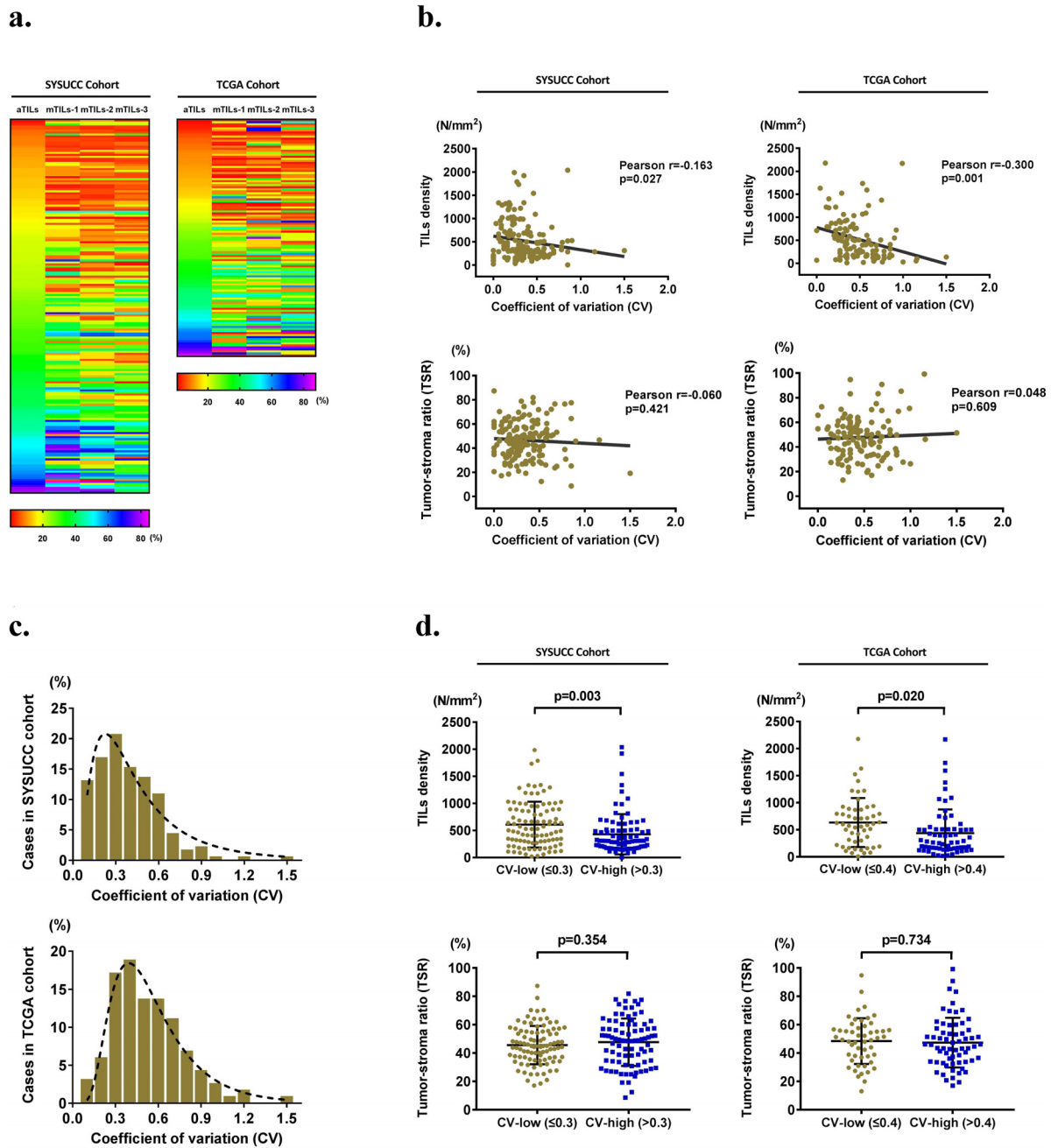


Fig. 2. The interobserver variation among VTAs. (a) Heatmap of the aTIL scores (as reference) and mTILs scores provided by three pathologists in both cohorts. (b) The CV was negatively correlated with TIL density, while there was no correlation between the CV and the TSR. (c) The distribution of CVs among the cases in both cohorts. (d) Differences in TIL density and TSR between the CV-low and CV-high cases in the SYSUCC cohort (CV cutoff at 0.3) and TCGA cohort (CV cutoff at 0.4).

(0.15–0.16) than in TCGA cohort (0.21–0.24). It is worth noted that mTILs and aTILs were not necessarily well calibrated although moderate correlation was found in between.

We subsequently assessed the agreement between mTIL and aTIL scores in the TILs-Low and TILs-High groups using four different cut-off points, 10%, 20%, 30%, and 40%, which have been reported in previous studies. We also generated a consensus group (mTILs-con) from individual dichotomized mTIL groups made by three pathologists on the basis of the majority voting rule. The kappa values of both cohorts are displayed in Fig. 4, and the values indicated a moderate-to-substantial agreement in groups determined using aTILs and those determined using individual/consensus mTILs scores at the cut-off points of 20% (kappa 0.36–0.58), 30% (kappa 0.36–0.50), and 40% (kappa 0.39–0.65) but not at 10% (kappa 0.15–0.21) in the SYSUCC

cohort. The optimal agreement between the mTIL and aTIL scores in terms of dichotomization was at the cutoff point of 40% in the SYSUCC cohort. The strength of agreement was lower in the TCGA cohort than in the SYSUCC cohort at all cutoff points (10%: kappa 0.15–0.28; 20%: kappa 0.33–0.46; 30%: kappa 0.30–0.33; 40%: kappa 0.21–0.31), and the optimal kappa value was observed at the cutoff point of 20% for the TIL score.

4.4. Prognostic values of mTILs and aTILs

In the SYSUCC cohort, the optimal cutoff points of the mTIL and aTIL scores for risk stratification were 9.0–15.0% (Supplementary Figure 2). DFS curves among TNBC cases in the TILs-Low and TILs-High groups with respect to the individual mTIL, mTILs-con, and aTIL

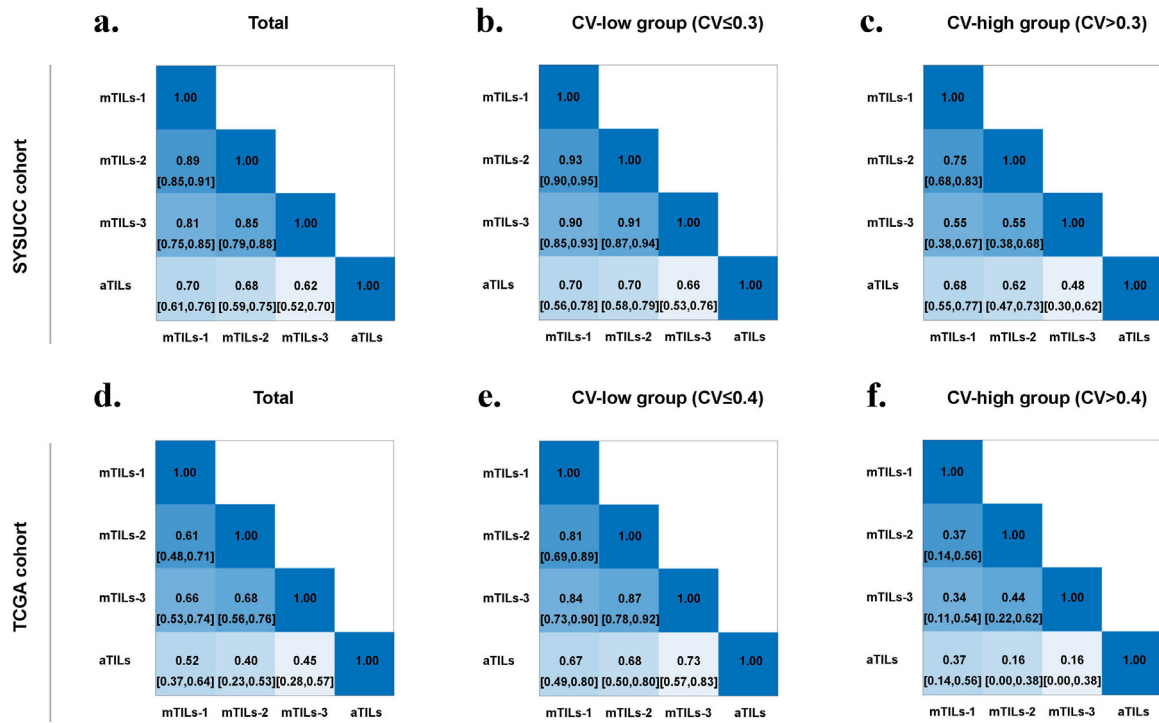


Fig. 3. Interobserver agreement among aTILs and mTILs scores. ICCs among the aTIL and mTIL scores of all the TNBC cases (a, d), as well as those in the CV-low (b, e) and CV-high groups (c, f). The cutoff values used in the SYSUCC and TCGA cohorts were 0.3 and 0.4, respectively.

scores are displayed in Fig. 5a-e, and the curves consistently show that TNBC cases in the TILs-High group have a higher DFS than do patients in the TILs-Low group. The HRs for the individual mTILs were 0.23 (mTILs-1; 95% CI 0.12–0.45, $p < 0.001$; log-rank test), 0.26 (mTILs-2; 95% CI 0.13–0.53, $p < 0.001$; log-rank test), and 0.34 (mTILs-3; 95% CI 0.19–0.64, $p < 0.001$; log-rank test). The HRs for mTILs-con and aTILs were 0.20 (95% CI 0.11–0.39, $p < 0.001$; log-rank test) and 0.30 (95% CI 0.16–0.59, $p < 0.001$; log-rank test), respectively. A Cox proportional hazards regression model was then used to identify the impact of stromal TILs, as either a continuous or

dichotomous variable, on the prognosis of patients with TNBC. The results are summarized in Table 2 and show that the stromal TIL density and individual mTIL, mTILs-con, and aTIL scores are considered significant and independent factors for DFS in univariate and multivariate analyses.

In the TCGA cohort, the optimal cutoff points for the mTIL and aTIL scores for risk stratification were 12.3–20.0% (Supplementary Figure 3). The survival analysis results shown in Fig. 5f-j revealed that the DFS in the TIL-High group was significantly higher than that in the TIL-Low group among TNBC cases after dichotomization by a

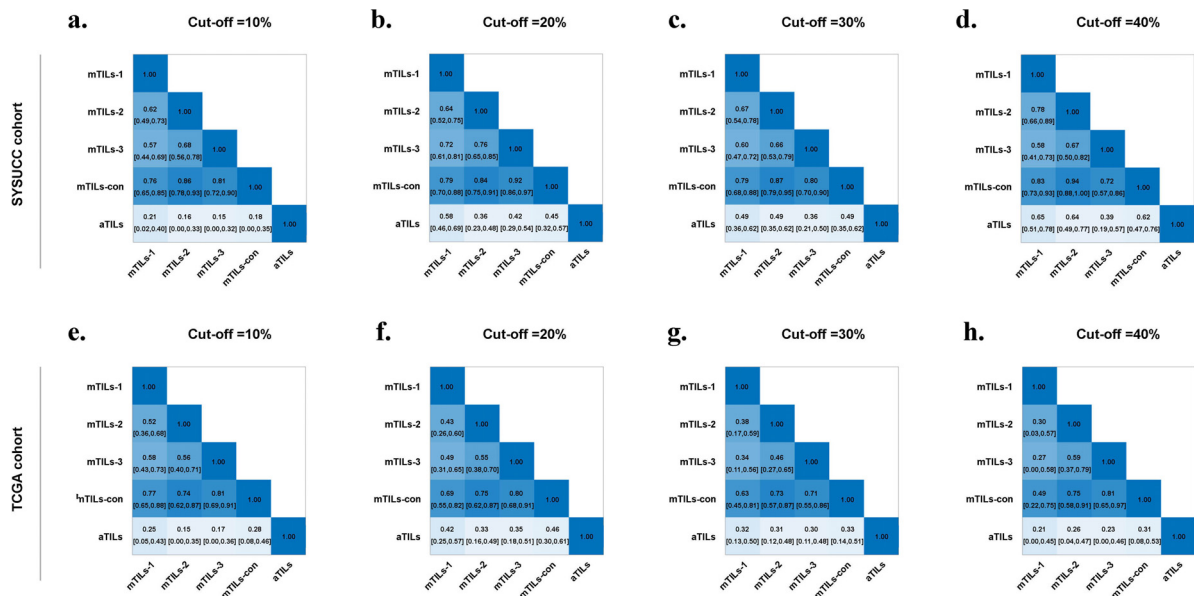


Fig. 4. Interobserver agreement among the aTIL, mTIL, and mTILs-con scores. The kappa values among aTILs, mTILs, and mTILs-con scores for the TILs-Low and TILs-High groups determined using a cutoff point of 10% (a, e), 20% (b, f), 30% (c, g), or 40% (d, h). In the mTILs-con model, a consensus group was generated from individual dichotomized mTIL scores grouped by three pathologists on the basis of the majority voting rule.

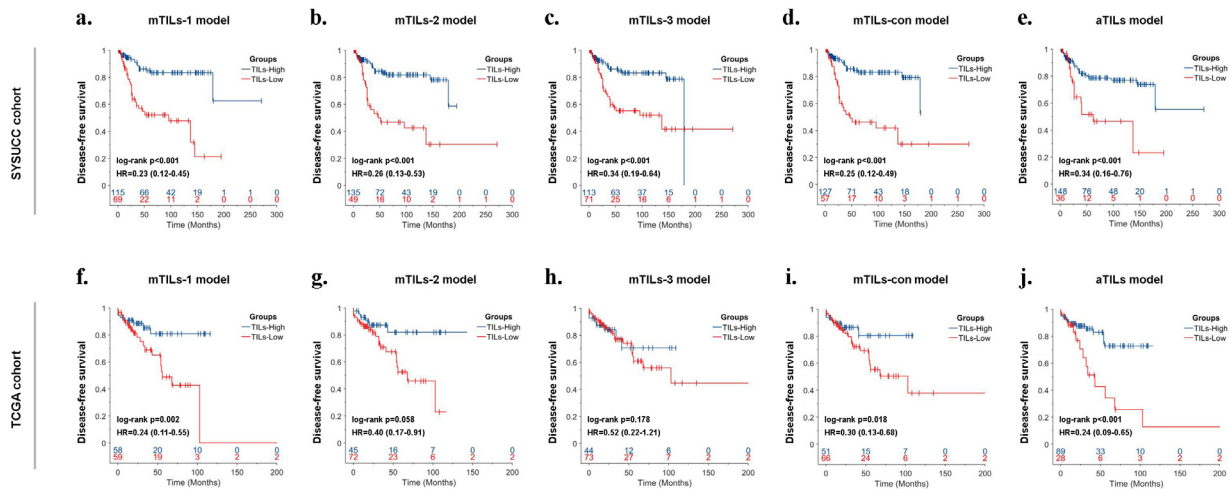


Fig. 5. DFS curves for TNBC cases using dichotomous models for stromal TILs. Kaplan–Meier curves of the DFS of TNBC patients in the TIL-Low and TIL-High groups based on dichotomization using mTIL-1 (a, f), mTIL-2 (b, g), mTIL-3 (c, h), mTIL-con (d, i), and aTIL models (e, j). The groups were established on the basis of the optimal cutoff stromal TIL scores for DFS using X-Tile. Details on the cutoff points used in the mTIL and aTIL models are demonstrated in Supplementary Figure 2 and Supplementary Figure 3. In the mTILs-con model, a consensus group was generated from individual dichotomized mTIL scores grouped by three pathologists based on the majority voting rule. The cut-off values used were 15%, 9%, 10% and 14.1% for mTILs-1, mTILs-2, mTILs-3 and aTILs in SYSUCC cohort and 15%, 20%, 20% and 12.3% for mTILs-1, mTILs-2, mTILs-3 and aTILs in TCGA cohort. P-value was calculated using log rank test.

senior pathologist by mTIL (mTILs-1: HR=0.24, 95% CI 0.11–0.55, $p = 0.002$; log-rank test), mTILs-con (HR=0.30, 95% CI 0.13–0.68, $p = 0.02$; log-rank test) and aTIL (HR=0.24, 95% CI 0.09–0.65, $p < 0.001$; log-rank test) scores. No significant differences were observed between groups after dichotomization by the mTILs scores provided by junior pathologists (mTILs-2: HR=0.40, 95% CI 0.17–0.91, $p = 0.058$; mTILs-3: HR=0.52, 95% CI 0.22–1.21, $p = 0.178$; log-rank test). For the continuous variables, the stromal TIL density, mTIL scores provided by a senior pathologist (mTILs-1), and aTIL scores were significantly associated with DFS in the univariate model. Only stromal TIL density and the aTIL scores were still considered significant and independent factors for DFS in further multivariate analyses (Table 2). For the categorical variables, mTILs-1, TILs-con, and aTILs were shown to be associated with DFS in the univariate model. However, only mTILs-1 and aTILs continued to be significant in further multivariate analyses (Table 2). The HRs of the mTIL and aTIL scores

determined using various cutoff points (10%, 20%, 30%, and 40%) in both cohorts are also shown in Supplementary Table 2.

4.5. Composite mTIL and aTIL model for prognostic stratification

On the basis of the different dichotomous outcomes using mTIL and aTIL scoring methods, we further stratified the cohort into three subgroups: the composite TILs-High group (mTILs-High/aTILs-High); the composite TILs-Low group (mTILs-Low/aTILs-Low); and the composite TILs-Uncertain group (either mTILs-High/aTILs-Low or mTILs-Low/aTILs-High). The DFS curves among groups in both cohorts are shown in Fig. 6. TNBC patients in the composite TILs-High group consistently showed the best DFS outcomes among the groups in the SYSUCC cohort, while the patients in the composite TILs-Low group showed the worst DFS among the groups in the TCGA cohort. The C index was calculated to measure the model's predictive performance

Table 2

Univariate and multivariate analyses of stromal TILs for DFS in SYSUCC and TCGA cohort.

Variable	SYSUCC cohort (N = 184)				TCGA cohort (N = 117)			
	Univariate model		Multivariate model ¹		Univariate model		Multivariate model ²	
	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value
Continuous variable								
TILs density (N/mm ²)	0.99 (0.99–1.00)	0.014	0.99 (0.99–1.00)	0.003	0.99 (0.99–1.00)	0.013	0.99 (0.99–1.00)	0.019
mTILs-1 (%)	0.96 (0.95–0.98)	<0.001	0.96 (0.94–0.98)	<0.001	0.97 (0.94–1.00)	0.028	0.97 (0.93–1.00)	0.067
mTILs-2 (%)	0.96 (0.94–0.98)	0.003	0.95 (0.93–0.98)	0.001	0.98 (0.96–1.00)	0.097	–	–
mTILs-3 (%)	0.96 (0.94–0.98)	<0.001	0.95 (0.92–0.98)	<0.001	0.99 (0.96–1.01)	0.230	–	–
aTILs (%)	0.97 (0.95–0.99)	0.009	0.97 (0.94–0.99)	0.016	0.96 (0.93–0.99)	0.005	0.96 (0.93–0.99)	0.007
Dichotomous variable								
mTILs-1 (high vs. low)	0.23 (0.12–0.42)	<0.001	0.19 (0.09–0.42)	<0.001	0.24 (0.09–0.64)	0.005	0.28 (0.10–0.79)	0.016
mTILs-2 (high vs. low)	0.26 (0.14–0.47)	<0.001	0.28 (0.14–0.55)	<0.001	0.40 (0.15–1.07)	0.068	–	–
mTILs-3 (high vs. low)	0.34 (0.18–0.62)	<0.001	0.24 (0.11–0.53)	<0.001	0.51 (0.19–1.39)	0.189	–	–
mTILs-con (high vs. low)	0.24 (0.13–0.44)	<0.001	0.23 (0.11–0.47)	<0.001	0.30 (0.10–0.87)	0.027	0.32 (0.11–0.91)	0.032
aTILs (high vs. low)	0.34 (0.18–0.63)	0.001	0.33 (0.16–0.68)	0.003	0.24 (0.11–0.54)	0.001	0.20 (0.08–0.50)	0.001

TILs, tumor infiltrating lymphocytes; DFS, disease-free survival; HR, hazard ratio; CI, confidence interval; mTILs, manual quantified stromal TILs score;

aTILs, automated quantified stromal TILs score; mTILs-con, a consensus group generated from individual dichotomized mTILs grouped by three pathologists based on majority voting rule. P-value was calculated using likelihood ratio test. The cut-off values used were 15%, 9%, 10% and 14.1% for mTILs-1, mTILs-2, mTILs-3 and aTILs in SYSUCC cohort and 15%, 20%, 20% and 12.3% for mTILs-1, mTILs-2, mTILs-3 and aTILs in TCGA cohort.

¹ Other variables included in the multivariate analysis were age, histological grade, tumor size, LNM(Y/N), TNM stage, chemotherapy(Y/N) and radiotherapy(Y/N).

² Other variables included in the multivariate analysis were age, histological grade, tumor size, LNM(Y/N) and TNM stage.

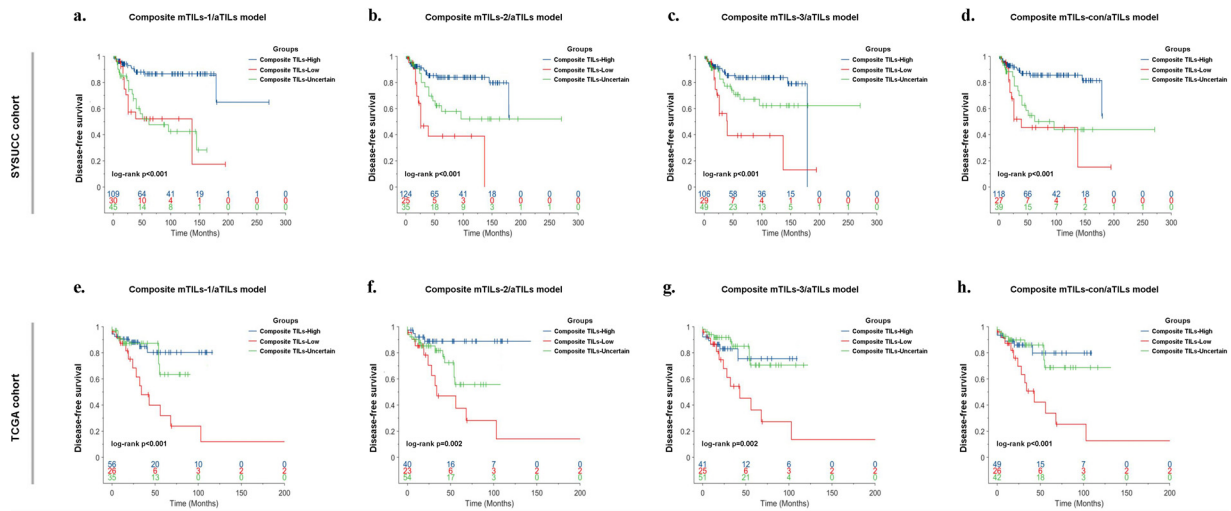


Fig. 6. DFS curves for TNBC cases using three-tier stratification models integrating aTILs and mTILs. TNBC cases in both cohorts were stratified into three subgroups based on the different dichotomous outcomes for the mTIL, mTILs-con, and aTIL models, including the composite TILs-High group (mTILs-High/aTILs-High), composite TILs-Low group (mTILs-Low/aTILs-Low), and composite TILs-Uncertain group (either mTILs-High/aTILs-Low or mTILs-Low/aTILs-High). Kaplan–Meier curves of DFS were generated for the three groups using the composite mTIL/aTIL models in the SYSUCC cohort (a–d) and TCGA cohort (e–h). The cut-off values used were 15%, 9%, 10% and 14.1% for mTILs-1, mTILs-2, mTILs-3 and aTILs in SYSUCC cohort and 15%, 20%, 20% and 12.3% for mTILs-1, mTILs-2, mTILs-3 and aTILs in TCGA cohort. P-value was calculated using log rank test.

for DFS with the mTIL, mTILs-con, and aTIL models, as well as the composite mTIL and aTIL models. We observed that the composite model demonstrated slightly higher C index ranges of 0.64 to 0.68 and 0.63 to 0.67 than did the previous two-tier stratification, which yielded C index ranges of 0.59 to 0.67 and 0.55 to 0.64 in the SYSUCC and TCGA cohorts, respectively (Table 3). The HRs of mTILs, mTILs-con, and aTILs in the three-tier stratification model in both cohorts are summarized in Table 4. The results revealed that for the three-tier stratification method, the composite mTIL and aTIL models were significantly associated with DFS in both the SYSUCC and TCGA cohorts. Notably, the mTIL scores provided by the two junior pathologists (mTILs-2 and mTILs-3), which were not found to be associated with DFS in the previous Cox regression analysis in the TCGA cohort, were significantly associated with DFS when composited with the aTIL scores in univariate (composite mTILs-2/aTILs, $p = 0.007$; composite mTILs-3/aTILs, $p = 0.005$; likelihood ratio test) and multivariate analyses (composite mTILs-2/aTILs, $p = 0.031$; composite mTILs-3/aTILs, $p = 0.033$; likelihood ratio test). Similarly, the TILs-con grouping method, which was not found to be significantly associated with DFS in the previous multivariate model, was considered an independent factor for DFS when the aTIL scores were composited in three-tier stratification.

In addition, the interclass comparison showed a significantly higher DFS in patients in the composite TIL-High group than in those in the composite TIL-Low group in both cohorts. In the SYSUCC cohort, compared with the composite TILs-Uncertain group, an increased DFS was observed in the composite TILs-Low group (univariate model: HR=0.40, 95% CI 0.18–0.88, $p = 0.023$; multivariate model: HR=0.42, 95% CI 0.18–0.96, $p = 0.040$; likelihood ratio test),

while a decreased DFS was found in the composite TILs-High group (univariate model: HR=2.40, 95% CI 1.18–4.92, $p = 0.016$; multivariate model: HR=3.01, 95% CI 1.41–6.40, $p = 0.004$; likelihood ratio test) with the composite mTILs-2/aTILs model. Similar survival outcomes were also observed among the composite mTILs-3/aTILs model according to the multivariate analysis (Table 4). However, with the mTILs-1/aTILs and mTILs-con/aTILs model, the patients in the composite TILs-Uncertain group were likely to have a DFS overlapping with those of the patients in the composite TILs-Low group. In contrast, the patients in the composite TILs-Uncertain group had a DFS similar to that of the patients in the composite TILs-Low group in the TCGA cohort.

5. Discussion

Stromal TILs have been verified to constitute a crucial prognostic and predictive factor for TNBC. Routine examinations of stromal TILs in patients with TNBC have also been recommended by the majority of experts who attended the St. Gallen International Breast Cancer Conference [34]. The International Immuno-Oncology Biomarker Working Group has developed a standardized assessment of TILs for breast cancer patients. Its value in terms of reproducibility and clinical use is highlighted by several previous studies [11–15], and both factors demonstrated acceptable interobserver agreement, with reported ICCs of 0.62–0.91. In the present study, similar ICCs (range, 0.61–0.93) were observed among three pathologists for the mTIL scores. However, VTAs still have inherent limitations, such as having interreader variability, having a risk of perceptual bias, and being time-consuming for comprehensive evaluations, which cannot be

Table 3
Harrell’s C-index of two-tier or three-tier stratification model using stromal TILs for DFS in SYSUCC and TCGA cohort.

Cohort	Two-tiers stratification model (c-index ± S.E)					Three-tiers stratification model (c-index ± S.E)			
	mTILs-1	mTILs-2	mTILs-3	mTILs-con	aTILs	mTILs-1/aTILs	mTILs-2/aTILs	mTILs-3/aTILs	mTILs-con/aTILs
SYSUCC	0.67 ± 0.04	0.64 ± 0.04	0.63 ± 0.04	0.66 ± 0.04	0.59 ± 0.04	0.68 ± 0.04	0.65 ± 0.04	0.64 ± 0.04	0.67 ± 0.04
TCGA	0.64 ± 0.05	0.57 ± 0.05	0.55 ± 0.05	0.61 ± 0.05	0.63 ± 0.06	0.67 ± 0.06	0.64 ± 0.06	0.63 ± 0.06	0.66 ± 0.06

TILs, tumor infiltrating lymphocytes; DFS, disease-free survival; mTILs, manual quantified stromal TILs score; aTILs, automated quantified stromal TILs score; mTILs-con, a consensus group generated from individual dichotomized mTILs grouped by three pathologists based on majority voting rule; S.E, standard error. The cut-off values used were 15%, 9%, 10% and 14.1% for mTILs-1, mTILs-2, mTILs-3 and aTILs in SYSUCC cohort and 15%, 20%, 20% and 12.3% for mTILs-1, mTILs-2, mTILs-3 and aTILs in TCGA cohort.

Table 4
Univariate and multivariate analyses in three-tier stratification model using stromal TILs for DFS in SYSUCC and TCGA cohort.

Variable	SYSUCC cohort (N = 184)				TCGA cohort (N = 117)			
	Univariate model		Multivariate model ¹		Univariate model		Multivariate model ²	
	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value	HR (95% CI)	p-value
composite mTILs-1/aTILs		<0.001		<0.001		0.001		0.001
uncertain vs. low	0.93 (0.45–1.90)	0.832	0.76 (0.34–1.70)	0.507	0.35 (0.13–0.94)	0.037	0.26 (0.09–0.77)	0.015
high vs. low	0.19 (0.09–0.41)	<0.001	0.14 (0.05–0.36)	<0.001	0.16 (0.06–0.44)	<0.001	0.14 (0.05–0.44)	0.001
uncertain vs. high	4.96 (2.41–10.19)	<0.001	5.57 (2.31–13.46)	<0.001	2.25 (0.68–7.43)	0.184	1.81 (0.53–6.15)	0.341
composite mTILs-2/aTILs		<0.001		0.001		0.007		0.008
uncertain vs. low	0.40 (0.18–0.88)	0.023	0.59 (0.24–1.41)	0.231	0.40 (0.17–0.98)	0.044	0.31 (0.11–0.81)	0.018
high vs. low	0.17 (0.08–0.35)	<0.001	0.20 (0.08–0.47)	<0.001	0.14 (0.04–0.51)	0.003	0.15 (0.04–0.57)	0.006
uncertain vs. high	2.40 (1.18–4.92)	0.016	2.95 (1.31–6.65)	0.009	2.89 (0.79–10.55)	0.107	2.09 (0.55–7.93)	0.277
composite mTILs-3/aTILs		<0.001		<0.001		0.005		0.001
uncertain vs. low	0.38 (0.18–0.81)	0.011	0.51 (0.22–1.18)	0.118	0.30 (0.12–0.74)	0.009	0.19 (0.07–0.53)	0.002
high vs. low	0.21 (0.10–0.43)	<0.001	0.19 (0.08–0.44)	<0.001	0.21 (0.07–0.66)	0.007	0.15 (0.05–0.51)	0.002
uncertain vs. high	1.82 (0.87–3.80)	0.110	2.78 (1.14–6.79)	0.025	1.42 (0.43–4.74)	0.566	1.23 (0.36–4.23)	0.745
composite mTILs-con/aTILs		<0.001		<0.001		0.002		0.001
uncertain vs. low	0.66 (0.31–1.39)	0.271	0.80 (0.34–1.85)	0.598	0.30 (0.12–0.77)	0.012	0.19 (0.07–0.53)	0.002
high vs. low	0.13 (0.06–0.30)	<0.001	0.18 (0.08–0.44)	<0.001	0.17 (0.05–0.51)	0.002	0.15 (0.05–0.51)	0.002
uncertain vs. high	3.52 (1.74–7.12)	<0.001	4.38 (1.91–10.03)	<0.001	1.83 (0.53–6.28)	0.336	1.23 (0.36–4.23)	0.745

TILs, tumor infiltrating lymphocytes; DFS, disease-free survival; HR, hazard ratio; CI, confidence interval; mTILs, manual quantified stromal TILs score.

aTILs, automated quantified stromal TILs score; mTILs-con, a consensus group generated from individual dichotomized mTILs grouped by three pathologists based on majority voting rule. P-value was calculated using likelihood ratio test. The cut-off values used were 15%, 9%, 10% and 14.1% for mTILs-1, mTILs-2, mTILs-3 and aTILs in SYSUCC cohort and 15%, 20%, 20% and 12.3% for mTILs-1, mTILs-2, mTILs-3 and aTILs in TCGA cohort.

¹ Other variables included in the multivariate analysis were age, histological grade, tumor size, LNM(Y/N), TNM stage, chemotherapy(Y/N) and radiotherapy(Y/N).

² Other variables included in the multivariate analysis were age, histological grade, tumor size, LNM(Y/N) and TNM stage.

fully resolved by standardization or training [18]. Our data also showed a higher interobserver variability in the TCGA cohort than in the SYSUCC cohort. Although the m-TIL score distributions were similar in the two cohorts, the pathologists declared that they are more familiar with the slides stained in their own center. The poor quality of the slides can explain why the pathologists had more difficulty with the TCGA dataset. Thick or faded sections, frozen sections, as well as those with scarce eosin staining or poor tissue fixation, are more commonly found in the TCGA cohort (examples are demonstrated in **Supplementary Figure 4**). Moreover, there are more outliers (determined as when the m-TIL scored by any pathologist deviated from that of the other two pathologists by over 20%) in the TCGA cohort than in the SYSUCC cohort (24/117, 20.5% vs. 20/184, 10.9%). The outliers more frequently occurred in cases with scarce stroma, abundant intratumoral TILs (**Supplementary Figure 5**), TILs surrounding residual lobules, sTILs mixed with neutrophils (**Supplementary Figure 6**), and extensive lymphovascular invasion (**Supplementary Figure 7**), which is consistent with the findings of Kos et al. [15] that interobserver variation was driven by heterogeneity in the lymphocyte distribution, staining procedure, tumor-associated stroma ratio, and tumor boundary definition, as well as lymphocytes being associated with other structures.

To overcome such limitations of VTAs, the current study aimed to develop a new CTA method and to investigate its clinical implications compared with VTAs. In a CTA workflow, the accurate identification and segmentation of intratumoral stromal regions and TILs, either in patches [23,35,36] or in individual TILs [37–39], are essential for stromal TIL analysis. We established a large annotated nuclei dataset to obtain a sufficient number of nuclei for classification training in our study. The accuracy of the nuclei classification model showed satisfactory performance for lymphocyte detection, and it was comparable to the level reported in the latest literature [40]. However, it should be noted that grouping all the nuclei other than malignant epithelial cells and TILs into “other” class was only feasible if the pathologist has demarcated tumor area beforehand, since some types of nuclei such as normal acini outside of the tumor area could be very similar to malignant epithelial cells. Intratumoral stroma area segmentation typically requires exhaustive boundary annotation, which is tedious and prone to high annotation errors. Most of the published

CTA studies [23,30,40,41] excluded this step. Amgad et al. [37] obtained a high-quality dataset with detailed area annotations through crowdsourcing and successfully trained an area segmentation model accordingly. In our study, the intratumoral stromal area was determined by the sliding window approach through a spatial distribution map of tumor cell nuclei. Compared with the segmentation model developed at the pixel level described by Amgad et al. [37,39], the segmentation mask generated by our method was relatively coarser, especially at the region boundaries. Thus, we asked a pathologist to inspect the stroma segmentation results and verify their suitability for subsequent analysis. In the future study, it would also be benefit if we can further solidate the necrosis and nuclei type ground truths by integrating annotations from various pathologists who are independent from mTILs scoring.

As suggested by the TILs-WG [9,18], the CTA algorithm should be developed in accordance with the guidelines for VTAs, and the fraction of intratumoral stroma occupied by TILs should be calculated. However, some of the existing CTA algorithms [23,38,40,41] did not recognize the stromal regions or identify individual TILs. The metrics established in these studies include lymphocyte hotspot-based detection [40,41] and TIL spatial distribution statistics [23,38]. To date, there has been no conclusion on which CTA algorithm is most appropriate. In a recent study by Le et al. [41], a patch-based CTA algorithm was adopted to calculate the ratio between lymphocyte patches and tumor patches. As Le et al. released their measurements for all BRCA cases in the TCGA cohort, we stratified those TNBC cases used in our study and found a similar prognostic performance between Le’s approach (HR=0.32, 95% CI 0.13–0.77, $p = 0.01$, log-rank test, C-index=0.634±0.06) and our approach (HR=0.24, 95% CI 0.09–0.65, $p < 0.001$, log-rank test, C-index=0.632±0.05), although our aTIL scores were determined in accordance with the manual scoring criteria to the greatest extent possible. These findings indicate the need for further validation of both types of approaches in the larger scale cohorts.

A few examples are presented to illustrate the details of consistent cases (**Supplementary Figures 8–10**) and inconsistent cases (**Supplementary Figures 11–13**) between the mTIL and aTIL scores. The possible reasons identified for the aTIL score to be inconsistent with the mTIL score include the following: 1.) the aTIL includes a possible

TLS region or dense lymphocyte clusters surrounding vessels, which are excluded by the pathologist in the mTIL (**Supplementary Figure 11**); 2.) the aTIL underestimated the stromal region when the tumoral stroma is sparse due to the resolution issue from the sliding window approach (**Supplementary Figure 12**); and 3.) aTIL tends to confuse the stromal area with loose appearance and necrotic areas (**Supplementary Figure 13**). Some of these discrepancies might be eliminated with improved algorithms, such as a finer stroma segmentation model and a specific TLS recognition model to exclude such areas. On the other hand, some of these issues, such as scarce stroma, TILs surrounding residual lobules, and extensive lymphovascular invasion, are also the causes for the interobserver variation between pathologists, as discussed previously. These difficulties might need to be more emphasized in the guidelines both for VTA and CTA. Another challenge we faced in developing CTA is the lack of quantitative criteria to determine how dense mononuclear infiltrates can be arranged and considered TILs-dense, especially if we wish to follow the VTA guideline. In our design, possible stroma areas were evaluated in $8\text{-}\mu\text{m}\times 8\text{-}\mu\text{m}$ (32×32 pixels) sliding windows, and those containing more than two mononuclear cells were considered a TIL-dense block. The choice of sliding window size and number of mononuclear cells inside each window were based on the overall consideration of image resolution and typical TIL diameters. The modification of this criteria might lead to a change in the aTIL scores, and further studies will be performed to systematically optimize the criteria in a larger cohort. Additionally, in the process of generating the aTIL score, we also calculated the overall TIL density and tumor-stroma ratio (TSR) across WSIs and evaluated their impact on the variation in mTIL scores. Our data suggested that the discordant VTA was more frequently observed in cases with low TIL density but was not relevant to TSR. These evidence might enable the identification of the elements causing ambiguity and interobserver variability of VTAs.

The CTA developed in this study could be used as an sTIL biomarker for TNBC prognosis as well as a complementary approach to VTA. The prognostic value of the CTA was verified in univariate and multivariate analyses in terms of DFS and was compared with VTA. We observed that the aTIL scores demonstrated better prognostic power than the mTIL scores, especially in the TCGA cohort. Nevertheless, an ideal cutoff point for stromal TILs is still desired for decision-making or risk management clinically, as in the ER, PR, HER2 workflows, as well as the recurrence rate (Oncotype DX); however, to date, no formal clinically relevant TILs cutoff points have been recommended. Previous VTA studies analyzing the prognostic impact of stromal TILs across various cutoff points on TNBC have yielded promising results [3–6,8]. A meta-analysis by Ibrahim et al. [42] also revealed that every 10% increment in TILs might lead to a 15–20% reduction in disease recurrence or mortality in TNBC patients. On the other hand, a cutoff point of 10% for the TIL fraction, which was used in the CTA algorithm developed by Amgad et al. [39], did not have any significant impact on overall survival in the breast cancer subgroup. In our study, we displayed the optimal cutoff of mTILs and aTILs scores for DFS, which varied between 9.0–20.0%. The prognostic power of aTILs was found to be significant under different optimal cutoff points in both cohorts. These findings suggested that our CTA model could be used individually in predicting prognoses under appropriate cutoff values.

On the other hand, inescapable variations were highlighted between mTILs and aTILs dichotomization. Kos et al. [15] mentioned that if only one singular cutoff point is used either for mTILs or aTILs, cases with the values around that cutoff point may frequently be misassigned. Thus, the present composite mTIL/aTIL model achieved a better prognostic performance than dichotomization using a single cutoff point, especially for junior pathologists. Over 73.4% and 52.7% of cases were assigned to either the composite TIL-High or composite TIL-Low group, respectively, which displayed a promising prognostic stratification result. This workflow complementing mTILs with aTILs

mitigates the overall risk of misassignment and enables pathologists to focus more on composite TIL-Uncertain group cases. Further analyses are subsequently required in "TILs-Uncertain" cases with the existing tools, including the reference images publicly available or online TNBC-prognosis tool [43] (www.tilsinbreastcancer.org), so that comprehensive, evidence-based decisions can be made for individuals [44]. Hence, we assume that our CTA model might be more suitable to be used as a computer-aided tool, which provided an on-site, immediate, second opinion regarding grouping according to the stromal TIL score for pathologists. An alternative workflow could be established under the computer-assisted setting to guide pathologist through the manual scoring process as suggested [18] for better standardization. Following studies will be required to compare both.

One strength of our study is that both VTA and CTA were assessed in Asian (SYSUCC) and Caucasian (TCGA) TNBC cohorts. As shown in **Supplementary Figure 14**, no significant differences were found for mTIL-2 ($p = 0.417$, student's t -test), mTIL-3 ($p = 0.751$, student's t -test) and aTIL ($p = 0.621$, student's t -test) scores between the two cohorts, while the mTIL-1 scores were slightly higher ($p = 0.004$, student's t -test) in the SYSUCC cohort. The optimized cut-off values of the two cohorts were the same for mTILs-1 (15% vs. 15%) but varied for mTILs-2 (9% vs. 20%), mTILs-3 (10% vs. 20%) and aTILs (14.1% vs. 12.3%). These findings suggested the possible need for different TILs cut-off values for different ethnicities, although no significant difference of TIL distribution was found between Asians and Caucasians. The prognosis performance was found slightly better in the SYSUCC cohort, but this result needs to be further validated as the cut-off values used were different in two cohorts.

In our CTA workflow, the invasive carcinoma boundary still needs to be manually delineated within all WSIs to reduce the impact of regional identification errors on the subsequent quantification results for stromal TILs. This process prevents our method from being fully automated. We should note that the identification of main tumor bulk and intratumoral stromal area, either manually or automatically, can be a challenge in some special histological subtypes of invasive breast carcinoma, such as lobular carcinomas, mucinous carcinomas, etc. In the present study, all cases included in SYSUCC cohort were diagnosed as invasive breast carcinoma of no specified type (IBC—NST). In TCGA cohort, 113 cases (96.6%) were IBC—NSTs, 3 cases (2.6%) were medullary carcinoma, which is currently categorized as TIL-rich IBC—NST or IBC—NST with medullary pattern. Only 1 case (0.8%) displayed pleomorphic lobular carcinoma in morphology with high nuclear grade. Thus, we need to clarify that the current CTA method is only validated on IBC—NSTs. Further tests and validations should be performed on other special histological subtypes of IBCs. From the technical point of view, automated region identification in breast cancer remains a challenging task, even with the most advanced deep learning techniques. In the recent ICIAR 2018 Breast Cancer Histology Image (BACH) Grand Challenge, the best pixel-level accuracy of the classification for normal epithelial cells, benign lesions, DCIS, and invasive carcinoma achieved from WSIs was lower than 70% [45]. Moreover, optimizations for our CTA algorithm are required to identify and handle other histological components including DCIS within invasive tumor, fibrosis, hyalinization, and larger number of granulocytes which are mentioned in VTA guideline.

In addition, the current CTA algorithm, as well as the composite mTIL/aTIL workflow, should be further verified in a larger multicenter cohort and integrated with standard clinicopathologic and genomic predictors to offer more practical insights. Independent validation in prognosis with identical cut-off values in both discovery and validation cohort should also be performed which is lack in the current study. Nonetheless, the CTA algorithm described in the present study provides a useful tool for stromal TIL assessments and prognosis evaluations in patients with TNBC.

6. Contributors

Conceptualization: PS, SX; Data acquisition: PS, XC, JH, ML, RL, SX; Methodology: KC, QH, JY; Data analysis: PS, XC, JH, SX, JK, HW, HL, HH; Writing original draft and editing: PS, SX, JH, XC; Data curation: PS, SX; Project administration: PS, SX; Funding acquisition: PS, SX. All authors read and approved the final manuscript.

Data sharing statement

The SYSUCC dataset used in this study are not publicly available due to institutional requirements governing privacy protection but could be available from the corresponding author upon reasonable request. The TCGA dataset used in this study were part of the TCGA-BRCA dataset which is publicly available through the Genomic Data Commons portal (<https://portal.gdc.cancer.gov/>).

Funding

National Natural Science Foundation of China, Guangdong Medical Research Foundation, Guangdong Natural Science Foundation.

Declaration of Competing Interest

SX and HH are cofounders of Bio-totem. JK and HL are fulltime employees at Bio-totem. HW was the intern at Bio-totem during this work. All remaining authors declare that no competing interests.

Acknowledgments

This work was funded by the Guangdong Medical Research Foundation (A2018241), the National Natural Science Foundation of China (81902679, 81700576), and the Guangdong Natural Science Foundation (2021A1515010760).

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.ebiom.2021.103492](https://doi.org/10.1016/j.ebiom.2021.103492).

References

- Loi S, Michiels S, Salgado R, Sirtaine N, Jose V, Fumagalli D, Kellokumpu-Lehtinen PL, Bono P, Kataja V, Desmedt C, Piccart MJ. Tumor-infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: results from the FinHER trial. *Ann. Oncol.* 2014 Aug 1;25(8):1544–50.
- Salgado R, Denkert C, Campbell C, Savas P, Nuciforo P, Aura C, De Azambuja E, Eidtmann H, Ellis CE, Baselga J, Piccart-Gebhart MJ. Tumor-infiltrating lymphocytes and associations with pathological complete response and event-free survival in HER2-positive early-stage breast cancer treated with lapatinib and trastuzumab: a secondary analysis of the NeoALTTO trial. *JAMA Oncol* 2015 Jul 1;1(4):448–55.
- Stanton SE, Disis ML. Clinical significance of tumor-infiltrating lymphocytes in breast cancer. *J Immunother Cancer* 2016 Dec;4(1):1–7.
- Stanton SE, Adams S, Disis ML. Variation in the incidence and magnitude of tumor-infiltrating lymphocytes in breast cancer subtypes: a systematic review. *JAMA Oncol* 2016 Oct 1;2(10):1354–60.
- Adams S, Gray RJ, Demaria S, Goldstein L, Perez EA, Shulman LN, Martino S, Wang M, Jones VE, Saphner TJ, Wolff AC. 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 Sep 20;32(27):2959.
- Pruneri G, Vingiani A, Bagnardi V, Rotmensz N, De Rose A, Palazzo A, Colleoni AM, Goldhirsch A, Viale G. Clinical validity of tumor-infiltrating lymphocytes analysis in patients with triple-negative breast cancer. *Ann. Oncol.* 2016 Feb 1;27(2):249–56.
- Luen SJ, Salgado R, Fox S, Savas P, Eng-Wong J, Clark E, Kiermaier A, Swain SM, Baselga J, Michiels S, Loi S. Tumour-infiltrating lymphocytes in advanced HER2-positive breast cancer treated with pertuzumab or placebo in addition to trastuzumab and docetaxel: a retrospective analysis of the CLEOPATRA study. *Lancet Oncol.* 2017 Jan 1;18(1):52–62.
- Savas P, Salgado R, Denkert C, Sotiriou C, Darcy PK, Smyth MJ, Loi S. Clinical relevance of host immunity in breast cancer: from TILs to the clinic. *Nat. Rev. Clin. Oncol.* 2016 Apr;13(4):228.
- Salgado R, Denkert C, Demaria S, Sirtaine N, Klauschen F, Pruneri G, Wienert S, Van den Eynden G, Baehner FL, Pénault-Llorca F, Perez EA. The evaluation of tumor-infiltrating lymphocytes (TILs) in breast cancer: recommendations by an International TILs Working Group 2014. *Ann. Oncol.* 2015 Feb 1;26(2):259–71.
- Hendry S, Salgado R, Gevaert T, Russell PA, John T, Thapa B, Christie M, Van De Vijver K, Estrada MV, Gonzalez-Ericsson PI, Sanders M. Assessing tumor-infiltrating lymphocytes in solid tumors: a practical review for pathologists and proposal for a standardized method from the International Immuno-Oncology Biomarkers Working Group: part 1: assessing the host immune response, TILs in invasive breast carcinoma and ductal carcinoma in situ, metastatic tumor deposits and areas for further research. *Adv Anat Pathol* 2017 Sep;24(5):235.
- Denkert C, Wienert S, Poterie A, Loibl S, Budczies J, Badve S, Bago-Horvath Z, Bane A, Bedri S, Brock J, Chmielik E. Standardized evaluation of tumor-infiltrating lymphocytes in breast cancer: results of the ring studies of the international immuno-oncology biomarker working group. *Mod. Pathol.* 2016 Oct;29(10):1155–64.
- Swisher SK, Wu Y, Castaneda CA, Lyons GR, Yang F, Tapia C, Wang X, Casavilca SA, Bassett R, Castillo M, Sahin A. Interobserver agreement between pathologists assessing tumor-infiltrating lymphocytes (TILs) in breast cancer using methodology proposed by the International TILs Working Group. *Ann. Surg. Oncol.* 2016 Jul;23(7):2242–8.
- Buisseret L, Desmedt C, Garaud S, Fornili M, Wang X, Van den Eyden G, de Wind A, Duquenne S, Boisson A, Naveaux C, Rothé F. Reliability of tumor-infiltrating lymphocyte and tertiary lymphoid structure assessment in human breast cancer. *Mod. Pathol.* 2017 Sep;30(9):1204–12.
- Khoury T, Peng X, Yan L, Wang D, Nagrale V. Tumor-infiltrating lymphocytes in breast cancer: evaluating interobserver variability, heterogeneity, and fidelity of scoring core biopsies. *Am. Clin. Pathol.* 2018 Oct 1;150(5):441–50.
- Kos Z, Roblin E, Kim RS, Michiels S, Gallas BD, Chen W, van de Vijver KK, Goel S, Adams S, Demaria S, Viale G. Pitfalls in assessing stromal tumor-infiltrating lymphocytes (STILs) in breast cancer. *NPJ breast cancer* 2020 May 12;6(1):1–6.
- Acs B, Rantalainen M, Hartman J. Artificial intelligence as the next step towards precision pathology. *J. Intern. Med.* 2020 Jul;288(1):62–81.
- Klauschen F, Müller KR, Binder A, Bockmayr M, Hägele M, Seegerer P, Wienert S, Pruneri G, De Maria S, Badve S, Michiels S. Scoring of tumor-infiltrating lymphocytes: from visual estimation to machine learning. *Seminars in cancer biology*, 52. Academic Press; 2018 Oct 1. p. 151–7.
- Amgad M, Stovgaard ES, Balslev E, Thagaard J, Chen W, Dudgeon S, Sharma A, Kerner JK, Denkert C, Yuan Y, Abdulljbar K. Report on computational assessment of tumor-infiltrating lymphocytes from the International Immuno-Oncology Biomarker Working group. *NPJ breast cancer* 2020 May 12;6(1):1–3.
- Bejnordi BE, Veta M, Van Diest PJ, Van Ginneken B, Karssemeijer N, Litjens G, Van Der Laak JA, Hermesen M, Manson QF, Balkenhol M, Geessink O. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017 Dec 12;318(22):2199–210.
- Bejnordi BE, Mullooly M, Pfeiffer RM, Fan S, Vacek PM, Weaver DL, Herschorn S, Brinton LA, van Ginneken B, Karssemeijer N, Beck AH. Using deep convolutional neural networks to identify and classify tumor-associated stroma in diagnostic breast biopsies. *Mod. Pathol.* 2018 Oct;31(10):1502–12.
- Jiménez G, Racoceanu D. Deep learning for semantic segmentation vs. classification in computational pathology: application to mitosis analysis in breast cancer grading. *Front Bioeng. Biotechnol.* 2019 Jun 21;7:145.
- Couture HD, Williams LA, Geradts J, Nyante SJ, Butler EN, Marron JS, Perou CM, Troester MA, Niethammer M. Image analysis with deep learning to predict breast cancer grade, ER status, histologic subtype, and intrinsic subtype. *NPJ breast cancer* 2018 Sep 3;4(1):1–8.
- Saltz J, Gupta R, Hou L, Kurc T, Singh P, Nguyen V, Samaras D, Shroyer KR, Zhao T, Batiste R, Van Arnam J. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell reports.* 2018 Apr 3;23(1):181–93.
- Abe N, Matsumoto H, Takamatsu R, Tamaki K, Takigami N, Uehara K, Kamada Y, Tamaki N, Motonari T, Unesoko M, Nakada N. Quantitative digital image analysis of tumor-infiltrating lymphocytes in HER2-positive breast cancer. *Virchows. Arch.* 2019 Dec 23;1–9.
- Hammond ME, Hayes DF, Dowsett M, Allred DC, Hagerty KL, Badve S, Fitzgibbons PL, Francis G, Goldstein NS, Hayes M, Hicks DG. American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). *Arch. Pathol. Lab. Med.* 2010 Jul;134(7):e48–72.
- Wolff ACEIzabeth, Hale Hammond M, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, et al. Human epidermal growth factor receptor 2 testing in breast cancer: american society of clinical oncology/college of American pathologists clinical practice guideline focused update. *J Clin Oncol* 2018;36(20) 2105–2.
- Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Kovatich AJ, Benz CC, Levine DA, Lee AV, Omberg L. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* 2018 Apr 5;173(2):400–16.
- Kumar N, Verma R, Anand D, Zhou Y, Onder OF, Tsougenis E, Chen H, Heng PA, Li J, Hu Z, Wang Y. A multi-organ nucleus segmentation challenge. *IEEE Trans Med Imaging* 2019 Oct 23;39(5):1380–91.
- He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. In: *Proceedings of the IEEE international conference on computer vision*; 2017. p. 2961–9.

- [30] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–8.
- [31] Sun P, Yeung PH, Xu S. Regional Convolutional Neural Network for Multi-Organ Nuclei Segmentation in H&E Pathology Images. LABORATORY investigation, Vol. 99. NEW YORK, NY 10013-1917 USA: NATURE PUBLISHING GROUP; 2019 Mar 1.
- [32] Chollet F. Xception: deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 1251–8.
- [33] Kassani SH, Kassani PH, Wesolowski MJ, Schneider KA, Deters R. Breast cancer diagnosis with transfer learning and global pooling. In 2019 International Conference on Information and Communication Technology Convergence (ICTC). IEEE; 2019 Oct 16. p. 519–24.
- [34] Burstein HJ, Curigliano G, Loibl S, Dubsy P, Gnani M, Poortmans P, Colleoni M, Denkert C, Piccart-Gebhart M, Regan M, Senn HJ. Estimating the benefits of therapy for early-stage breast cancer: the St. Gallen International Consensus Guidelines for the primary therapy of early breast cancer 2019. *Ann. Oncol.* 2019 Oct 1;30(10):1541–57.
- [35] Amgad M, Elfandy H, Hussein H, Atteya LA, Elsebaie MA, Abo Elnasr LS, Sakr RA, Salem HS, Ismail AF, Saad AM, Ahmed J. Structured crowdsourcing enables convolutional segmentation of histology images. *Bioinformatics.* 2019 Sep 15;35(18):3461–7.
- [36] Basavanthally AN, Ganesan S, Agner S, Monaco JP, Feldman MD, Tomaszewski JE, Bhanot G, Madabhushi A. Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology. *IEEE Trans. Biomed. Eng.* 2009 Oct 30;57(3):642–53.
- [37] Amgad M, Sarkar A, Srinivas C, Redman R, Ratra S, Bechert CJ, Calhoun BC, Mrazek K, Kurkure U, Cooper LA, Barnes M. Joint region and nucleus segmentation for characterization of tumor infiltrating lymphocytes in breast cancer. *Medical imaging 2019: digital pathology*, 10956. International Society for Optics and Photonics; 2019 Mar 18:109560M.
- [38] Yuan Y, Failmezger H, Rueda OM, Ali HR, Gräf S, Chin SF, Schwarz RF, Curtis C, Dunning MJ, Bardwell H, Johnson N. Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Sci Transl Med* 2012 Oct 24;4(157) 157ra143-.
- [39] Heindl A, Sestak I, Naidoo K, Cuzick J, Dowsett M, Yuan Y. Relevance of spatial heterogeneity of immune infiltration for predicting risk of recurrence after endocrine therapy of ER+ breast cancer. *J. Natl. Cancer Inst.* 2018 Feb 1;110(2):166–75.
- [40] Sirinukunwattana K, Raza SE, Tsang YW, Snead DR, Cree IA, Rajpoot NM. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Trans Med Imaging* 2016 Feb 4;35(5):1196–206.
- [41] Le H, Gupta R, Hou L, Abousamra S, Fassler D, Torre-Healy L, Moffitt RA, Kurc T, Samaras D, Batiste R, Zhao T. Utilizing automated breast cancer detection to identify spatial distributions of tumor-infiltrating lymphocytes in invasive breast cancer. *Am. J. Pathol.* 2020 Jul 1;190(7):1491–504.
- [42] Ibrahim EM, Al-Foheidi ME, Al-Mansour MM, Kazkaz GA. The prognostic value of tumor-infiltrating lymphocytes in triple-negative breast cancer: a meta-analysis. *Breast Cancer Res. Treat.* 2014 Dec;148(3):467–76.
- [43] Loi S, Drubay D, Adams S, Pruneri G, Francis PA, Lacroix-Triki M, Joensuu H, Dieci MV, Badve S, Demaria S, Gray R. Tumor-infiltrating lymphocytes and prognosis: a pooled individual patient analysis of early-stage triple-negative breast cancers. *J. Clin. Oncol.* 2019 Mar 1;37(7):559.
- [44] Hipp J, Flotte T, Monaco J, Cheng J, Madabhushi A, Yagi Y, Rodriguez-Canales J, Emmert-Buck M, Dugan MC, Hewitt S, Toner M. Computer aided diagnostic tools aim to empower rather than replace pathologists: lessons learned from computational chess. *J Pathol Inform* 2011;2.
- [45] Aresta G, Araújo T, Kwok S, Chennamsetty SS, Safwan M, Alex V, Marami B, Prastawa M, Chan M, Donovan M, Bach Fernandez G. Grand challenge on breast cancer histology images. *Med Image Anal* 2019 Aug 1;56:122–39.