

# Supplementary Material

## A ML models fine-tuning and hyperparameters

	Cohort B (1 times 4-fold)	Cohort B (5 times 4-fold)	Cohort A (5 times 4-fold)
MeanPool	<b>0.6091 ± 0.0486</b>	<b>0.5661 ± 0.0544</b>	<b>0.7135 ± 0.0779</b>
Weldon	0.5587 ± 0.0600	0.5553 ± 0.0520	0.6681 ± 0.1047
Chowder	0.5881 ± 0.0240	0.5479 ± 0.0500	0.5896 ± 0.1438
DeepMIL	0.5827 ± 0.0300	0.5420 ± 0.0612	0.6775 ± 0.0656

**Table 1.** Mean AUCs reached by selected hyperparameters from one 4-fold cross validation grid search on Cohort B.

Model	Hyperparameters Grid
CART	$D = [1, 2, 3, 4, 5, 6, 7, 8]$
XGBoost	$D = [1, 2, 3, 4, 5, 6, 7, 8], \eta = [0.001, 0.005, 0.01, 0.05, 0.1, 0.5]$
ElasticNet	$C = [1., 10., 100., 1000., 0.1, 0.01, 0.001]$

**Table 2.** Hyper-parameters grid

Indicator		Cohort A	Cohort B
# slides	Training	82	367
	Test	20	60
	Total	102	427
Pathology Data		Bouin Fixative until 2010 then Formol TILs annotated by one pathologist a posteriori	AFA Fixative before 2016 then Formol TILs annotated by multiple pathologists through the years, reflecting clinical practice
Specific clinical data		Sum of Tumor Sizes, Principal Tumor Sizes, pM	KI67, FISH, insitu carcinoma, mitosis, necrosis
Clinical data status		No missing data	Some missing data (9.22% on average on common variables)

Table 3: Overview of the cohorts

## B Cohorts and variables description

Clinical Feature	Description	Unit	Availability
<i>Patient Information</i>			
Sex	Sex of the patient	M F	A/B
Weight	Patient weight at time of diagnosis.	kg	A/B
Age	Patient age according to one of four pre-specified age ranges.	(0-25] (25- 50] (50- 75] (75- 100]	A/B
Body-Mass Index (BMI)	Patient BMI at the time of diagnosis.	kg / m <sup>2</sup>	A/B
Menopause Status	Patient Menopause status at the time of diagnosis.	pre post unspecified	A/B
<i>Pre-Treatment Measurements</i>			
Biopsy Year	The year at which the diagnosing biopsy was collected.	year	A/B
Side	On which breast was the diagnosis made.	left right	A/B
Insitu Carcinoma	Whether the biopsy contains insitu carcinoma.	Yes No	B
Multifocal Tumor	The cancer is multifocal when there are at least two separated lesions.	Yes No	A/B
cT	PRE-treatment clinical TNM status. Categorization of primary tumor.	Tx/Na T0 Tis T1 T2 T3 T4a T4b T4c T4d	A/B
cN	PRE-treatment clinical TNM status. Categorization of spread to regional lymph nodes.	Nx/No N1 N2 N3	A/B
cN	PRE-treatment clinical TNM status. Presence of distant metastasis.	M0 M1	A/B
<i>Pre-Treatment Pathology Information (pathologist assessment)</i>			

<b>TILs</b>	Percentage of tumor-infiltrating T-lymphocytes observed in the pre-treatment histology; given as a continuous value. Assessed by one or multiple pathologists, depending on the cohort.	[0-100] percentage	A/B
<b>EE grade</b>	Elston and Ellis Histological grade.	1 2 3	A/B
<b>Histo. Subtype</b>	Classification of which of the three primary breast cancer subtypes the diagnosis belongs to.	Invasive Carcinoma ISC- Ductal ISC-Lobular	A/B
<b>PR</b>	The progesterone-receptor-positive (PR+) status of the tumor, graded using IHC.	Uninterpretable Negative or (0, 10%] Negative or (10, 50%] Negative or (50, 80%] Negative or (80, 100%]	A/B
<b>ER</b>	The estrogen-receptor-positive (PR+) status of the tumor, graded using IHC.	Uninterpretable Negative or (0, 10%] Negative or (10, 50%] Negative or (50, 80%] Negative or (80, 100%]	A/B A/B
<b>HER2-IHC</b>	Result of Immunohistochemistry-based test for HER2 in pathology. A value of +++ represents a confident HER2 positive measurement, while ++ is borderline, and 0/+ represents an HER2-negative test.	0/+ ++ +++	A/B
<b>FISH</b>	Fluorescence In Situ Hybridization (FISH), a more accurate measurement which can be employed to corroborate the IHC-based HER2 test for the HER2 status of tumors cells.	Amplification + Absence of Amplification -	B
<b>KI67</b>	Ki-67 Labelling Index. Measurement of the presence of Antigen Ki-67 (Marker of Proliferation, Ki-67), a nuclear protein associated with cellular proliferation.	[0-100] percentage	B
<b>Presence of Vascular Emboli</b>		Present Not Present	A/B
<b>Necrose</b>	Whether or not necrotic tissue was observed in pathology.	Observed Not Observed	
<b>Mitosis</b>		1 2 3	B
<b><i>Post-Treatment Measurements</i></b>			
<b>pT</b>	POST-treatment pathological TNM status. Categorization of primary tumor	Tx/Na T0 Tis T1 T2 T3	A/B

		T4a T4b T4c T4d	
<b>pN</b>	POST-treatment pathological TNM status. Categorization of spread to regional lymph nodes.	Tx/Na Nx/No N1 N2 N3	A/B
<b>pM</b>	POST-treatment pathological TNM status. Presence of distant metastasis.	M0 M1	A
<b>RCB</b>	Combination of measurements of tumor, tumor bed, and metastasis present in axillary lymph nodes	continuous index	B
<b>RCB Class</b>	Severity scoring of the continuous RCB measurement. Scoring is calculated according to predetermined thresholds (0, 1.36, 3.28).	0 I II III	B
<b>Principal Tumor Size</b>	Direct measurement of the principal tumor after treatment.	mm	A
<b>Sum of Tumor Sizes</b>	Sum of sizes of all identified tumors after treatment.	mm	A
<b>NACT Histological Response</b>	<b>Study Endpoint.</b> Classification of whether or not the patient responded to neo-adjuvant chemotherapy: thresholding of RCB Class.	RCB Class 0 – Complete RCB Class I II III – Incomplete	A/B

Table 4: Description and categorization of available per-patient clinical data across all cohorts. Cells in yellow indicate admission measurements which serve as criteria for admission into the study cohorts.

## C Analysis of differences in Cohorts A and B

We compare the distribution of each common clinical variable between both cohorts in order to track the origin of the differences in the predictions’ performances. We display Fig. 5 the univariate histograms of the most discriminative variable between the two cohorts: age, EE grade and cT status. By inspecting the oldest group of patients in Cohort B, we see that among 14 patients only 1 had a positive complete response to NACT, in the rest of the cohort there are 51.2% good responders. This result is expected because older patients are more at risks.

These comparisons of clinical data distributions between the two cohorts demonstrate that despite the alignment of data between the two cohorts, the cohorts are still heterogeneous. In addition, we note that in Cohort B, biopsies before 2017 are from a retrospective cohort, whereas biopsies received after 2017 were not curated. However, the original source of disparities between both populations is still relatively unknown as it does not seem to appear strongly in the distributions of the extracted clinical variables.

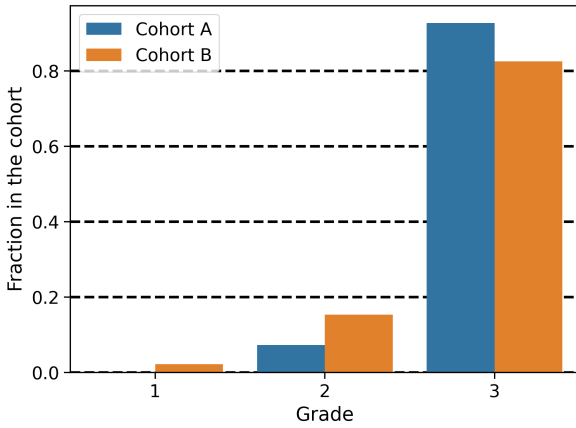
## D TNBC studies patient count

## E Clinical Practice: Model’s Confusion Analysis

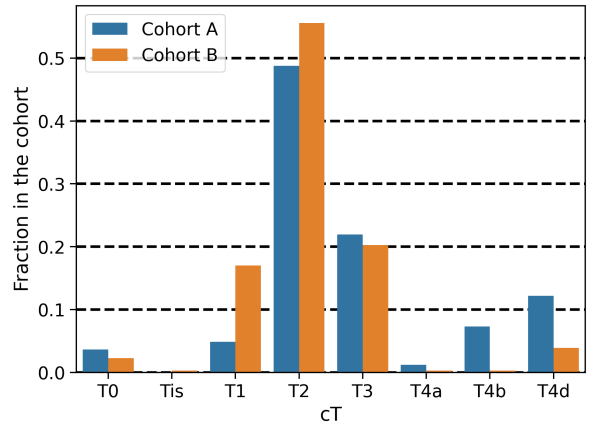
For each trained model, we complement Fig. 3 b) by providing confusion matrices, for the best model of each type in each cohort, by optimizing the decision threshold in order to maximise the specificity while keeping a maximum of 25% false negative patients.

## F Annotation Pathology Criteria

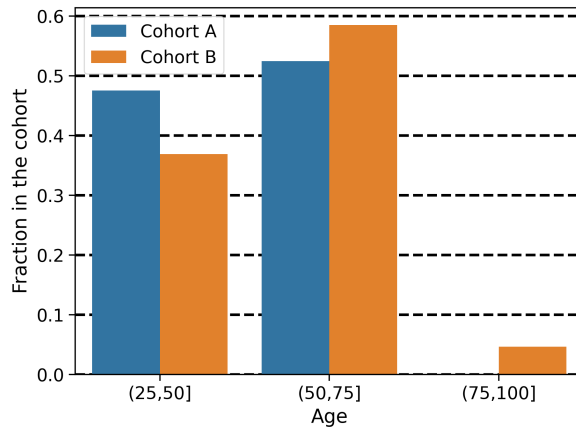
In Table 5, each criterion is associated with the values it can take whether it is percentage, discrete levels (grading or presence/absence of specific levels), or boolean (presence/absence). Scoring and annotations of the architecture, atypia and mitosis in the first rows of the table refer to the 3 components of the EE tumor



(a) EE grade distribution



(b) cT distribution

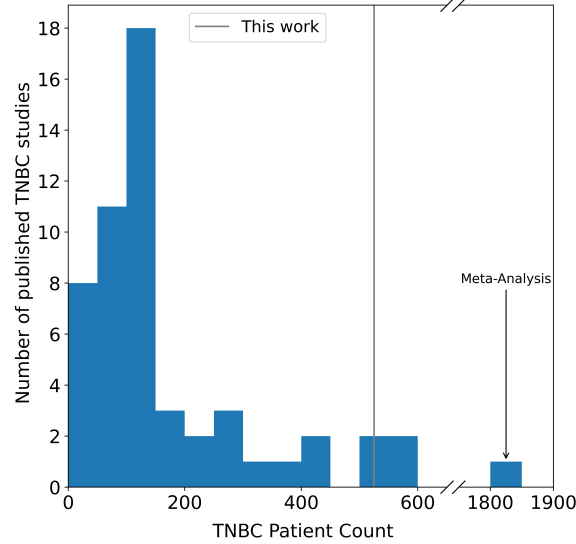


(c) Age distribution

**Fig. 5.** Distribution of the most discriminative variables between the two cohorts: EE grade, cT and age in each cohort

grade. Regarding the Dominant architecture annotations, we call fibrosis the results from the deposition of a cross-linked collagen matrix by cancer-associated fibroblasts (CAFs). We also call discohesive tumor cells, the classic histological micro pattern of ILC (Invasive Lobular Carcinoma) defined as "characterized by proliferation of non-cohesive small cells individually dispersed in fibrous connective tissue or organized in single-file linear cords invading the stroma" [88].

**Fig. 6.** TNBC studies patient counts.

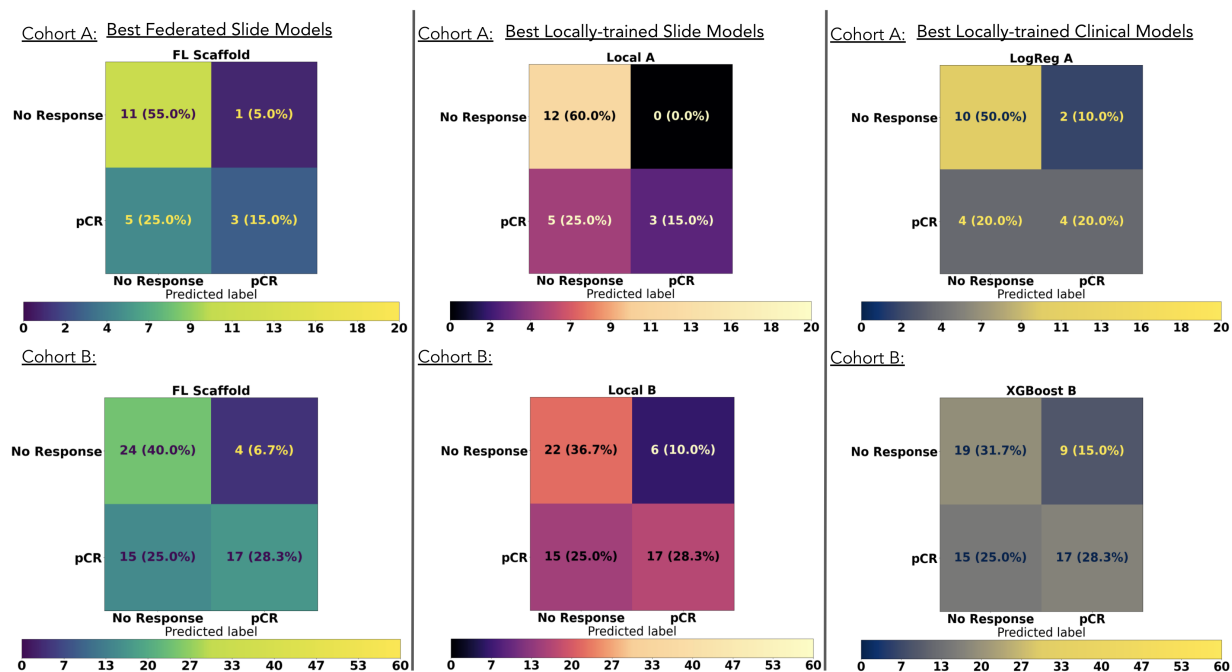


Patient count in recent TNBC studies following a comprehensive bibliographic search. We extracted 54 articles from either the survey from Bianchini et al. [86] or from our article using all citations that contain the keywords "triple negative", "triple-negative" or "TNBC" and from which the number of patients could be extracted upon reading. Out of 55 TNBC studies including ours, our article is in the top 10% in terms of patient count. Median patient count is 119, with 25-75% quantiles 82-207. The study on the far right from Couch et al. [87] with 1,800 patients is a meta-analysis of multiple other studies.

<b>Epithelioid Structures</b>	Tumor cellularity	presence/absence
	Architecture	1 to 3
	Dominant Architecture	solid, span, discohesive cells, cribriform, tubes, micropapillary
	Atypias	1 to 3
	Mitosis	presence/absence
	Dominant Aspect of the Tumor cells visible nucleoli	pleomorphic, fusiform, monomorphic, cytoplasm
	Abundance of Cytoplasm	Yes, No
	Mucus	Yes, No
	Necrosis	presence/absence
	Tumorous Emboli	presence/absence
	Epidural sheathing	presence/absence
	<b>Stromal Structures</b>	Stroma
TILs		% of the tile area
Other inflammatory cells		neutrophils, macrophages, nothing
Fibrosis		presence/absence
Elastose		presence/absence
Calcifications		presence/absence
<b>Normal Structures</b>	Normal Glands	presence/absence
	Vascular Structures	presence/absence
	Hemosiderin	presence/absence
	Hemorrhagic Suffusion	presence/absence
	Mantle tissue (adipocytes, fibrosis)	presence/absence
<b>Artifacts</b>	detachment, out of focus, wrinkles, No artefact	Yes, No

Table 5: Tiles Pathology Criteria

Fig. 7. Confusion Analysis of models.



We display on the sanctuarized test sets, the confusion matrices obtained by choosing the classification threshold maximizing specificity while producing in each Cohort A maximum of 25% false negatives patients. For clinical practice false negatives is the worst outcome possible: patients that could be cured by the standard treatment are oriented towards second line treatments potentially more toxic and hazardous. For each model category we choose the best model according to its AUC point estimate.