European Journal of Cancer 143 (2021) 134–146

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Original Research

Identification of breast cancer patients with pathologic complete response in the breast after neoadjuvant systemic treatment by an intelligent vacuum-assisted biopsy☆,☆☆

André Pfob [a], Chris Sidey-Gibbons [b], Han-Byoel Lee [c,d,e], Marios Konstantinos Tasoulis [f], Vivian Koelbel [a], Michael Golatta [a], Gaiane M. Rauch [g], Benjamin D. Smith [h], Vicente Valero [i], Wonshik Han [c,d,e], Fiona MacNeill [f], Walter Paul Weber [j], Geraldine Rauch [k], Henry M. Kuerer [l], Joerg Heil [a,*]

[a] *Department of Gynecology, Heidelberg University, Heidelberg, Germany*
[b] *Department of Symptom Research, The University of Texas MD Anderson Cancer Center, Houston, USA*
[c] *Department of Surgery, Seoul National University College of Medicine, Seoul, South Korea*
[d] *Biomedical Research Institute, Seoul National University Hospital, Seoul, South Korea*
[e] *Cancer Research Institute, Seoul National University, Seoul, South Korea*
[f] *Department of Breast Surgery, The Royal Marsden NHS Foundation Trust, London, United Kingdom*
[g] *Department of Diagnostic Radiology, The University of Texas MD Anderson Cancer Center, Houston, USA*
[h] *Department of Radiation Oncology, University of Texas MD Anderson Cancer Center, Houston, USA*
[i] *Department of Breast Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, USA*
[j] *Department of Breast Surgery, University Hospital Basel and University of Basel, Basel, Switzerland*
[k] *Institute of Biometry and Clinical Epidemiology, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, And Berlin Institute of Health, Berlin, Germany*
[l] *Department of Breast Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, USA*

Received 6 September 2020; received in revised form 26 October 2020; accepted 9 November 2020
Available online 8 December 2020

☆ Final results were presented as a late breaking abstract (P2-42) and as part of an oral educational session "De-escalation of Surgical Therapy: What Does the Data Support - Exceptional Responders to Neoadjuvant Chemotherapy" (ES5-3) at the San Antonio Breast Cancer Symposium 2020, December 8–12, 2020.

☆☆ Preliminary results were presented at the American Society of Clinical Oncology (ASCO) 2020 annual meeting on May 29th, 2020 (abstract number 565).

* *Corresponding author*. Department of Gynecology, Heidelberg University, Im Neuenheimer Feld 440 69120 Heidelberg, Germany.Fax: +49 6221 5633681
E-mail address: Joerg.Heil@med.uni-heidelberg.de (J. Heil).

https://doi.org/10.1016/j.ejca.2020.11.006
0959-8049/© 2020 Elsevier Ltd. All rights reserved.

KEYWORDS
Pathologic complete response;
Neoadjuvant systemic treatment;
Breast cancer;
Individualized treatment;
Machine learning;
Artificial intelligence;
Vacuum-assisted biopsy;
Surgical oncology

Abstract ***Background:*** Neoadjuvant systemic treatment elicits a pathologic complete response (pCR) in about 35% of women with breast cancer. In such cases, breast surgery may be considered overtreatment. We evaluated multivariate algorithms using patient, tumor, and vacuum-assisted biopsy (VAB) variables to identify patients with breast pCR.
Methods: We developed and tested four multivariate algorithms: a logistic regression with elastic net penalty, an Extreme Gradient Boosting (XGBoost) tree, Support Vector Machines (SVM), and neural network. We used data from 457 women, randomly partitioned into training and test set (2:1), enrolled in three trials with stage 1–3 breast cancer, undergoing VAB before surgery. False-negative rate (FNR) and specificity were the main outcome measures. The best performing algorithm was validated in an independent fourth trial.
Results: In the test set (n = 152), the logistic regression with elastic net penalty, XGboost tree, SVM, and neural network revealed an FNR of 1.2% (1 of 85 patients with missed residual cancer). Specificity of the logistic regression with elastic net penalty was 52.2% (35 of 67 women with surgically confirmed breast pCR identified), of the XGBoost tree 55.2% (37 of 67), of SVM 62.7% (42 of 67), and of the neural network 67.2% (45 of 67). External validation (n = 50) of the neural network showed an FNR of 0% (0 of 27) and a specificity of 65.2% (15 of 23). Area under the ROC curve for the neural network was 0.97 (95% CI, 0.94–1.00).
Conclusion: A multivariate algorithm can accurately select breast cancer patients without residual cancer after neoadjuvant treatment.
© 2020 Elsevier Ltd. All rights reserved.

Abbreviations

VAB	image-guided, vacuum-assisted biopsy
pCR	pathologic complete response
ROC	receiver operating characteristic
CI	confidence interval
SD	standard deviation
SLNB	sentinel lymph node biopsy
ALND	axillary lymph node dissection
NST	neoadjuvant systemic treatment
SHAP	Shapley additive explanations
XGBoost	Extreme Gradient Boosting
LIME	Local Interpretable Model-Agnostic Explanations

1. Background

Approximately 20–40% of patients with primary breast cancer undergo neoadjuvant chemotherapy with or without anti-HER2 treatment [1]. Depending on tumor biology and stage, about 35% receiving neoadjuvant systemic treatment (NST) achieve a pathologic complete response (pCR) in the breast; novel treatment regimens have shown pCR rates of up to 80% in patients with triple-negative cancers and up to 70% in HER-2 positive cancers [2–5]. For these women without residual disease after NST, subsequent breast-conserving surgery or even mastectomy may have no therapeutic effect as all tumor cells have already been eradicated by NST. However, breast surgery is currently still required for these patients to diagnose whether or not residual disease is left after NST. An alternative, less invasive approach to identify women without residual disease after NST may help reducing morbidity by reducing potentially unnecessary surgical interventions. In the era of multimodality management, reduction of surgical interventions could reduce the burden of treatment for the patient, provider, and funder. For example, primary breast-ablative surgery has been de-escalated to breast-conserving surgery [6,7] and primary axillary lymph node dissection (ALND) has been de-escalated to sentinel lymph node biopsy (SLNB) in the adjuvant treatment setting [8,9]. These are well-evidenced examples that allowed safe de-escalation on the basis of modern multimodality diagnostics and treatment.

However, approaches to confirm pCR in the breast (ypT0) without surgery have revealed predominately negative results: the diagnostic accuracy of imaging (ultrasonography, mammography, magnetic resonance imaging, and positron emission tomography–computed tomography) is insufficient to confirm breast pCR [10,11]. Some single-center trials using vacuum-assisted biopsy (VAB) to confirm breast pCR yielded promising results [12,13]—but recent subsequent confirmatory, multicenter trials could not reach their primary endpoint: compared to standard breast surgery, VAB showed high false-negative rates ranging from 18% to 50% meaning that in many patients with tumor in the surgical specimen there was no residual tumor in the VAB specimen [14–17].

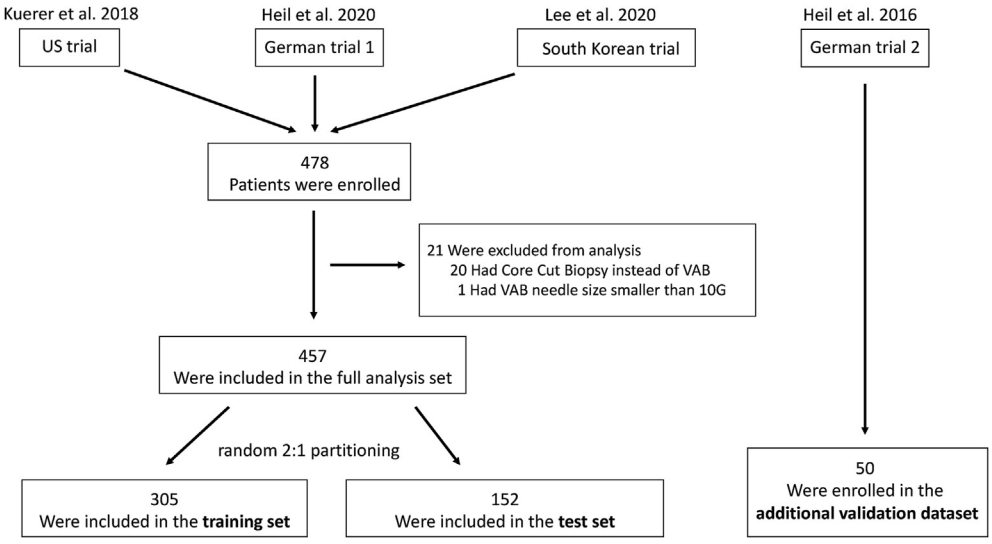

Fig. 1. Study ddesign.

One of these recent multicenter VAB trials showed in an exploratory analysis that combining the results of VAB and imaging resulted in relevantly less missed residual cancer [14]. Thus, one approach that may help overcome the challenge to accurately predict breast pCR is the use of multivariate algorithms which can simultaneously incorporate tumor, patient, imaging, and VAB variables—multivariate algorithms might allow a more individualized and accurate prediction of breast pCR.

In this diagnostic study, we aimed to develop and test different multivariate algorithms (logistic regression with elastic net penalty, Extreme Gradient Boosting (XGBoost) tree model, Support Vector Machines (SVM), neural network) using patient, tumor, and VAB variables ("intelligent VAB") to identify patients with breast pCR after neoadjuvant systemic treatment.

2. Methods

2.1. Patient recruitment and selection

We analyzed patients who participated in prospective studies assessing the feasibility of VAB to accurately detect residual disease after NST [12–14,18]. The studies were conducted at 23 sites in the United States, Germany, and South Korea. The trials enrolled women who presented with clinical stage I–III breast cancer of any biological subtype and had a partial or complete response to NST confirmed by ultrasonography, mammography, or magnetic resonance imaging. Evaluation of response on imaging was conducted as part of clinical routine according to standard guidelines to ensure generalizability [19,20]. The minimal invasive VAB procedure was performed before surgery, under either ultrasonography or stereotactic guidance.

The full analysis set of the training, test, and additional validation set was composed of all women who met the inclusion and exclusion criteria for the respective trial and underwent image-guided VAB with needles ranging from 10G to 7G (Fig. 1).

All trials and their combined anonymized analysis were approved by the respective institutional review boards and ethics committees. All participants gave written informed consent to participate in the respective trials.

2.2. Outcomes and definitions

Histopathologic evaluation of disease response in the surgical specimen was the reference standard. All patients underwent guideline-adherent surgery after NST and the histopathologic evaluation was conducted according to standard guidelines [19]. Pathologic complete response in the breast was defined as the absence of residual invasive and *in situ* tumor cells in the surgical specimen (ypT0).

The false-negative rate (FNR) was defined as all findings of no residual tumor by the index test (see below: imaging, VAB, the combination of VAB and imaging, the multivariate algorithms) among all cases with residual tumor on final pathologic evaluation of the surgical specimen (reference test).

We used specificity to quantify possible future reduction of surgery. Specificity was defined as all cases with breast pCR by the index test divided by all cases with a breast pCR on final pathologic evaluation of the surgical specimen (reference test).

2.3. Algorithm development

We developed and tested four algorithms with increasing complexity. Choice of algorithms and reporting on them was informed by guidelines on how to use machine learning in medicine [21], how to report

Table 1
Variables used for algorithm development.

Variable	Classification[a]	Definition
age	numerical (years)	numerical age of patient
imaging before NST	numerical (mm)	largest diameter as assessed by either ultrasound and mammography and/or MRI as applicable in clinical routine
imaging after NST	numerical (mm)	largest diameter as assessed by either ultrasound and mammography and/or MRI as applicable in clinical routine
multifocal disease on imaging	either (1) before NST (2) after NST (3) none	more than one lesion in one breast quadrant
multicentric disease on imaging	either (1) before NST (2) after NST (3) none	lesions in more than one breast quadrant
tumor grading	either (1) G1 (2) G2 (3) G3 (4) Gx	according to ASCO/CAP and German S3 guidelines, evaluated by board certified pathologist on the initial diagnostic biopsy (not on VAB or surgery specimen)
tumor biology	Either (1) No Special Type (NST) (2) Invasive Lobular Carcinoma (ILC) (3) Other	according to ASCO/CAP and German S3 guidelines, evaluated by board certified pathologist on the initial diagnostic biopsy (not on VAB or surgery specimen)
additional *in situ* carcinoma	yes/no	according to ASCO/CAP and German S3 guidelines, evaluated by board certified pathologist on the initial diagnostic biopsy (not on VAB or surgery specimen)
estrogen receptor status	positive/negative	according to ASCO/CAP and German S3 guidelines, positive if $\geq 10\%$[b]
progesterone receptor status	positive/negative	according to ASCO/CAP and German S3 guidelines, positive if $\geq 10\%$[b]
Her2Neu receptor status	positive/negative	according to ASCO/CAP and German S3 guidelines, positive if status ≥ 3 or amplified by FISH/CISH
Ki67 score	numerical score	according to ASCO/CAP and German S3 guidelines, hot spot evaluation
Neoadjuvant Systemic Treatment regimens	yes/no for each (1) Anthracycline (2) Texane (3) Platinum (4) Trastuzumab (5) Pertuzumab (6) other	Neoadjuvant Systemic Treatment regimens used
position of Clip marker to tumor lesion after NST	Either (1) Present and within the (former) lesion (2) Present and ≤ 5 mm distance to the (former) lesion (3) Present and >5 mm distance to the (former) lesion (4) Not present	position of Clip marker to the tumor lesion on imaging after NST
needle size used for VAB	Either (1) 10G (2) 9G (3) 8G (4) 7G	needle size used for VAB
guidance used for VAB procedure	Either (1) sonography (2) stereotaxy	guidance used for VAB procedure
representative VAB according to biopsying physician	yes/no	subjective assessment of the biopsying physician
number of specimens taken during VAB procedure	numerical	number of cores taken
pathologic evaluation of VAB sample	either (1)–(3); yes/no for (4) (1) tumor cells in VAB sample (*in situ* or invasive) (2) no tumor cells and VAB sample representative of former tumor region (visible signs of former tumor region like fibrosis)	evaluated by board certified pathologist, independently evaluated from surgical specimen

(*continued on next page*)

Table 1 (*continued*)

Variable	Classification[a]	Definition
	(3) no tumor cells and VAB sample unrepresentative of former tumor region (no visible signs of former tumor region like fibrosis) (4) difficulties with pathologic evaluation of VAB sample (subjective assessment of responsible pathologist)	

NST = neoadjuvant systemic treatment, VAB = vacuum-assisted biopsy.
[a] Dichotomized if applicable for analysis.
[b] We are aware that a new group of low estrogen receptor positive tumors (1–9%) is currently discussed. As research found these tumors to behave similar to hormone receptor negative tumors [54] these tumors are considered as hormone receptor negative.

findings of diagnostic tests [22] and multivariate prediction models [23], as well as previously published research by our group [24–26]. We provide a detailed description of all algorithms and the algorithm development as well as a detailed evaluation of our study according to these guidelines [21–23] in the online Supplementary Appendix.

The four approaches were:

1) Logistic regression with elastic net penalty [27,28].
2) XGBoost tree model [29,30]. The Shapley additive explanations (SHAP) method was used to provide insights into the black-box model. SHAP assigns an importance value for a particular prediction to each variable and thus allows an interpretation which variables were most important in making a certain prediction [31].
3) SVM with radial basis function kernel [32,33].
4) Deep, multi-layer neural network [34] with rectified Linear Unit activation [35,36] and Adam optimizer with Nesterov momentum model optimization [37]. The Local Interpretable Model-Agnostic Explanations (LIME) method was used to provide insights into the black-box model. LIME values reflect the contribution of each variable to the prediction of a data sample by applying local, linear approximations to the underlying, complex model [38].

Anonymized data of three prospective trials [13,14,18] were randomly partitioned into a training set and a test set (2:1). The three trials used for the algorithm development and testing reported FNRs (missed residual disease) for VAB of 17.8% (patients analyzed n = 398) [14], 30.8% (n = 40) [18], and 5% (n = 40) [13].

The performance of the four algorithms was evaluated with respect to FNR and specificity on the test set. Because the output of all four multivariate algorithms is a continuous risk probability, a single threshold (operating point) for each algorithm was determined to calculate and report FNR, specificity, NPV, and PPV. This threshold was determined as the maximum probability which resulted in a maximum of one false-negative case on the test set. Histograms plotting the model predictions against the actual outcomes were used to assess not only the classification but also the calibration of the algorithms. Also area-under-the-receiver-operating-characteristic-curve (AUC) was used to evaluate algorithm performance.

The best performing algorithm was then validated using another, fourth trial. This trial reported a false-negative rate for VAB of 25.9% (patients analyzed n = 50) [12].

Point estimates of the outcomes along with two-sided 95% Clopper-Pearson confidence intervals (CIs) are provided.

We used the PROBAST tool [39] to assess the risk of bias and the applicability of our model.

2.4. Role of the funding source

There was no funding source for this analysis and manuscript. The single trials were supported by a Cancer Center Support Grant from the National Institutes of Health (NIH) (CA16672), a NIH P30 grant (CA016672), and funding from the MD Anderson Clinical Research Funding Award Program [13], the German German Research Foundation (DFG) (GZ:HE 6824/5-1) [14], the Seoul National University Hospital Research Fund (30-2016-0250), and the Institute for Information and Communications Technology Promotion (IITP) grant funded by the Korea government (2018-0-00861) [18].

3. Results

3.1. Patients and datasets

A total of 478 women were included in the three international studies described above. Twenty-one patients were excluded because they had core cut biopsy instead of VAB or the biopsy needle was smaller than 10G; the remaining 457 patients in the full analysis set were randomly partitioned (2:1) into a training set (n = 305) and a test set (n = 152). Patients from a fourth trial were used as an additional validation dataset for the most promising algorithm (n = 50). We used 19 patient, tumor, and VAB variables for our predictive models (see Table 1 for definitions).

With respect to the distribution of baseline data between the four trials, diameter size on imaging before NST was lowest in the German trial 1 (mean size 25.2 mm) and highest in the German trial 2 (mean size 31.8 mm, $p = 0.008$); diameter size on imaging after

NST was lowest in the German trial 1 (mean size 7.1 mm) and highest in the South Korean trial (mean size 10.9 mm, $p = 0.023$). Significant differences were observed for the distribution of triple-negative and luminal tumors ($p = 0.009$, $p = 0.002$, respectively; lower numbers in the German trials 1 and 2 compared the US trial and the South Korean trial). HER2-positive tumors and tumor biology (NST, ILC, others) were equally distributed between the four trials ($p = 0.657$, $p = 0.507$, respectively).

3.2. Baseline demographic and clinical characteristics

Baseline patient demographic and clinical characteristics of the 457 patients in the full analysis set are summarized in Table 2.

Examination of the breast surgical specimen (reference test) showed overall breast pCR (ypT0 status) in 217 patients (47.5%). By tumor biological subtype, breast pCR was detected in 103 of 177 patients (58.2%) with triple-negative breast cancer, 35 of 50 (70.0%) with HER2-positive cancer/hormone receptor negative, 46 of 104 (44.2%) with HER2-positive cancer/hormone receptor positive cancer, 1 of 12 (8.3%) with luminal A, and 32 of 114 (28.1%) with Luminal B-like HER2-negative cancer.

3.3. Performance of the four multivariate algorithms and of imaging and/or VAB

Table 3 shows the diagnostic performance of VAB or routine clinical imaging response assessment alone and their combination (assume breast pCR when no tumor on imaging AND no tumor in VAB) in the detection of residual disease after NST, as well as the performance of the multivariate algorithms. All displayed analyses were evaluated on the same test set (n = 152). Detailed cross tabulations are provided in the online Supplementary Appendix.

The calibration of the deep neural network was better compared to the other three multivariate algorithms (Fig. 3).

The one false-negative case of the four multivariate algorithms (out of 85 patients with residual tumor in the test set) showed residual hormone receptor positive, HER2 negative ypT1b, ypN0 disease with tumor cellularity <30% (Table 3).

3.4. Insights into variable importance of two black-box models

Fig. 2 provides insights into the variable importance of the two black-box models Extreme Gradient Boosting tree (Fig. 2a, Shapley additive explanations (SHAP) value summary plot) and Deep Neural Network (Fig. 2b, LIME summary plot) by using local interpretation methods.

4. Discussion

In this diagnostic study we developed and tested four multivariate models ("intelligent VAB") to accurately identify patients with breast pCR after neoadjuvant systemic treatment. All four models showed a relevantly lower false-negative rate compared to previous clinical

Table 2
Baseline demographic and clinical characteristics of participating women.

Characteristic	Value
Mean age (SD)—yr	51.84 (11.69)
Age—no. (%)	
<30 yr	10 (2.2)
30–50 yr	198 (43.3)
51–70 yr	219 (47.9)
>70 yr	30 (6.6)
Grade—no. (%)	
1	5 (1.1)
2	147 (32.2)
3	301 (65.9)
x	4 (0.9)
Tumor biology—no. (%)	
Her2neu-positive/hormone receptor negative	50 (10.9)
Her2neu-positive/hormone receptor positive	104 (22.8)
Triple-negative	177 (38.7)
Luminal A-like	12 (2.6)
Luminal B-like	114 (24.9)
cT category—no. (%)	
cT0	0 (0.0)
cT1a	1 (0.2)
cT1b	37 (8.1)
cT1c	160 (35.0)
cT2	232 (50.8)
cT3	27 (5.9)
cT4	0 (0.0)
ycT category—no. (%)	
ycT0	191 (41.8)
ycT1a	34 (7.4)
ycT1b	98 (21.4)
ycT1c	99 (21.7)
ycT2	32 (7.0)
ycT3	3 (0.7)
ypT category—no. (%)	
ypT0	217 (47.5)
ypT1a	64 (14.0)
ypT1b	38 (8.3)
ypT1c	53 (11.6)
ypT2	41 (9.0)
ypT3	4 (0.9)
ypT4	0 (0.0)
ypTis	40 (8.8)
ypN category—no. (%)	
ypN0	342 (74.8)
ypN+	69 (15.1)
ypNx	46 (10.1)
ypT0 stratified for ypN—no. (%)	
ypT0 and ypN0	185 (85.2)
ypT0 and ypN+	9 (4.1)
ypT0 and ypNx	23 (10.6)
Findings on examination of VAB specimen—no. (%)	
Residual tumor	181 (39.6)
No residual tumor and representative	233 (51.0)
No residual tumor and not representative	43 (9.4)

VAB = vacuum-assisted biopsy.

Table 3
Performance testing of the multivariate algorithms compared to standard imaging and vacuum-assisted biopsy.

	False-negative rate—% (95% CI); no.	Specificity—% (95% CI); no.	Negative predictive value—% (95% CI); no.	Positive predictive value—% (95% CI); no.	AUROC—value (95% CI)
Test set (n = 152)					
Imaging	25.9% (17.0–36.5%); 22 of 85	61.2% (48.5–72.9%); 41 of 67	65.1% (52.0–76.7%); 41 of 63	70.8% (60.2–79.9%); 63 of 89	–
VAB	16.5% (9.3–26.1%); 14 of 85	89.6% (79.7–95.7%); 60 of 67	81.1% (70.3–89.3%); 60 of 74	91.0% (82.4–96.3%); 71 of 78	–
Imaging + VAB	5.9% (1.9–13.2%); 5 of 85	52.2% (39.7–64.6%); 35 of 67	87.5% (73.2–95.8%); 35 of 40	71.4% (62.1–79.6%); 80 of 112	–
Logistic regression with elastic net penalty	1.2% (0.0–6.4%); 1 of 85	52.2% (39.7–64.6%); 35 of 67	97.2% (85.5–99.9%); 35 of 36	72.4% (63.3–80.3%); 84 of 116	0.97 (0.95–1.00)
Extreme Gradient Boosting tree	1.2% (0.0–6.4%); 1 of 85	55.2% (42.6–67.4%); 37 of 85	97.4% (86.2–99.9%); 37 of 38	73.7% (64.6–81.5%); 84 of 114	0.97 (0.95–0.99)
Support Vector Machine	1.2% (0.0–6.4%); 1 of 85	62.7% (50.0 –74.2%); 42 of 67	97.7% (87.7–99.9%); 42 of 43	77.1% (68.0–84.6%); 84 of 109	0.97 (0.94–0.99)
Deep Neural Network	1.2% (0.0–6.4%); 1 of 85	67.2% (54.6 –78.2%); 45 of 67	97.8% (88.5–99.9%); 45 of 46	79.3% (70.3–86.5%); 84 of 106	0.97 (0.95–0.99)
Validation set (n = 50)					
Deep Neural Network	0.0% (0.0–12.8%); 0 of 27	65.2% (42.7–83.6%); 14 of 23	100% (78.2–100%); 15 of 15	77.1% (59.9–89.6%); 27 of 35	0.93 (0.86–1.00)

AUROC = Area-under-the-receiver operating characteristic curve; CI = confidence interval.

decision tools like imaging and/or VAB. Specificity of all four algorithms was comparable to imaging but inferior to VAB. As hypothesized, the performance among the four models improved with increasing complexity (logistic regression with elastic net penalty, XGBoost tree, SupportSVMs, neural network). FNR (missed residual disease) of the deep neural network was very low in the multicenter test sample (false-negative rate of 1%) as well as in an additional validation cohort (FNR of 0%). Specificity (ability to identify breast pCR) was 67% in the test set and 65% in the additional validation cohort.

To our knowledge, this concept of an intelligent VAB is the first evidence that the majority of patients with breast pCR could be accurately identified with a very low FNR to miss residual disease. After recent multicenter trials reporting negative results on this topic [14–17], our present analysis re-opens the discussion about the future management of exceptional responders to NST: a prospective diagnostic trial to confirm our findings or even a prospective therapeutic trial to carefully evaluate long-term oncologic outcomes of omitting breast surgery for women with algorithm-diagnosed breast pCR after NST seem warranted.

In interpreting and applying the findings of our analysis, some issues need to be addressed.

First, the value of further de-escalation of breast oncologic surgery is highly controversial. Lumpectomy is a procedure with very low morbidity compared to other oncological surgeries, but the treatment burden seems still relevant to patients: among patients undergoing low-morbidity breast-conserving surgery and SLNB, 40–50% experience persistent pain [40] and patients report a relevant reduction of patient-reported quality of life [41]. Although morbidity is also attributed to radiation treatment, the reduction of surgery-associated treatment burden in the past (radical mastectomy, to simple mastectomy, to lumpectomy) showed a positive impact on patient-reported quality of life and oncologic safety could be ensured [42].

Second, for a novel, less invasive diagnostic tool like our intelligent VAB, a very low rate of missed cancer (FNR) compared to the previous gold standard (breast surgery) is the most important measure to warrant subsequent therapeutic trials comparing long-term oncologic outcomes of the two approaches [43]—especially in times of post-neoadjuvant treatment regimes [42,44,45]. The exact multimodal management protocol for a possible future therapeutic trial will have to be carefully chosen and incorporate clinicians' and patients' considerations [46,47]. Although a low FNR is clinically most relevant, specificity is of importance, too. Our results show that multivariate algorithms can achieve a very low FNR (1%) compared to imaging (26%) and VAB (17%), but specificity can be further improved: specificity of the four multivariate algorithms (52%–

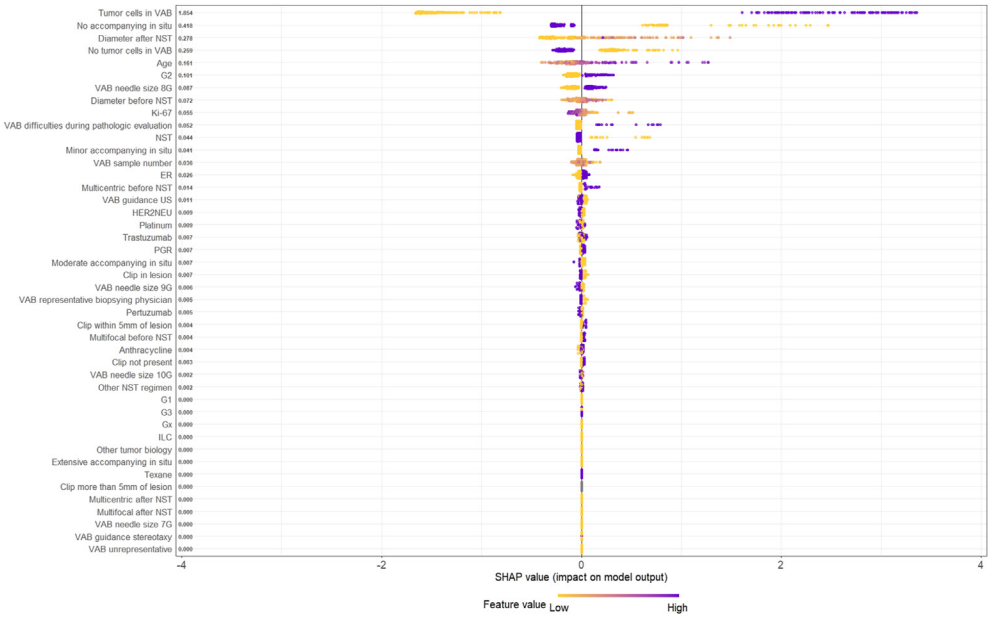

Fig. 2. Insights into variable importance of the Extreme Gradient Boosting tree model and the Deep Neural Network using local interpretation methods. (a) Shapley additive explanations (SHAP) value summary plot of the Extreme Gradient Boosting (XGBoost) tree model. Positive SHAP values on the x-axis indicate that the variable was important for predicting residual tumor in the breast; negative values indicates that the variable was important for predicting no residual tumor in the breast. Purple indicates a high variable value (e.g., tumor cells in VAB: yes); yellow indicates a low variable value (e.g., tumor cells in VAB: no). The values on the y-axis represent the overall global variable importance. (b) Local Interpretable Model-Agnostic Explanations (LIME) summary plot for the Deep Neural Network and its predictions on the test set. Blue indicates that the variable was important for predicting residual tumor in the breast; red indicates that the variable was important for predicting no residual tumor in the breast. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

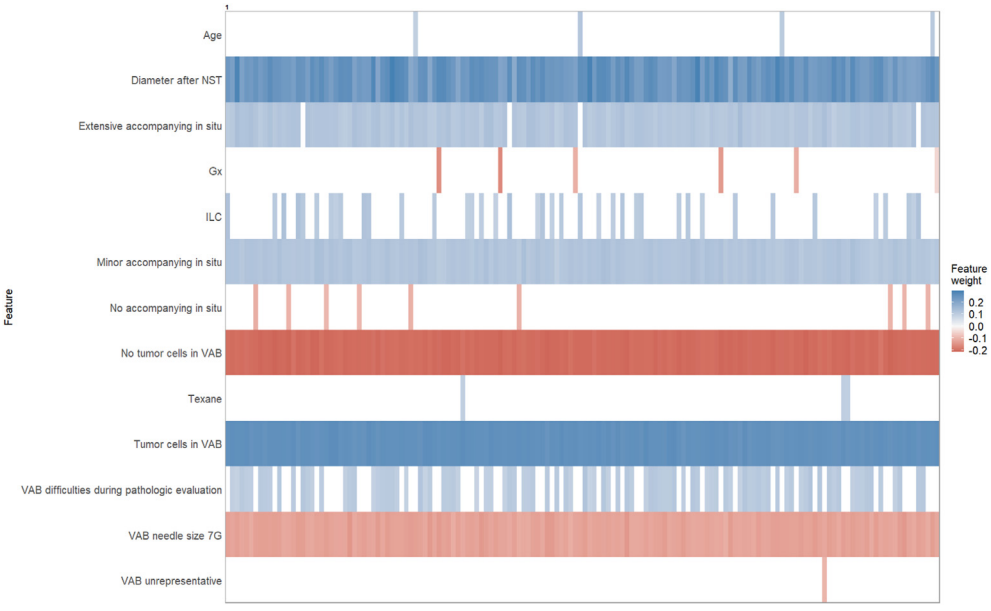

Fig. 2. (*continued*).

67%) was comparable to imaging (61%) but inferior to VAB (89%) alone.

Third, the exact patient population eligible for further surgical de-escalation in breast cancer treatment is yet to be defined. With respect to tumor biology, patients with triple-negative and HER-2 positive tumor show the highest rates of pCR (60% and 45% in our sample) whereas especially Luminal A tumors rarely achieve pCR (8% in our sample). Axillary status is also important for defining a future patient population: our

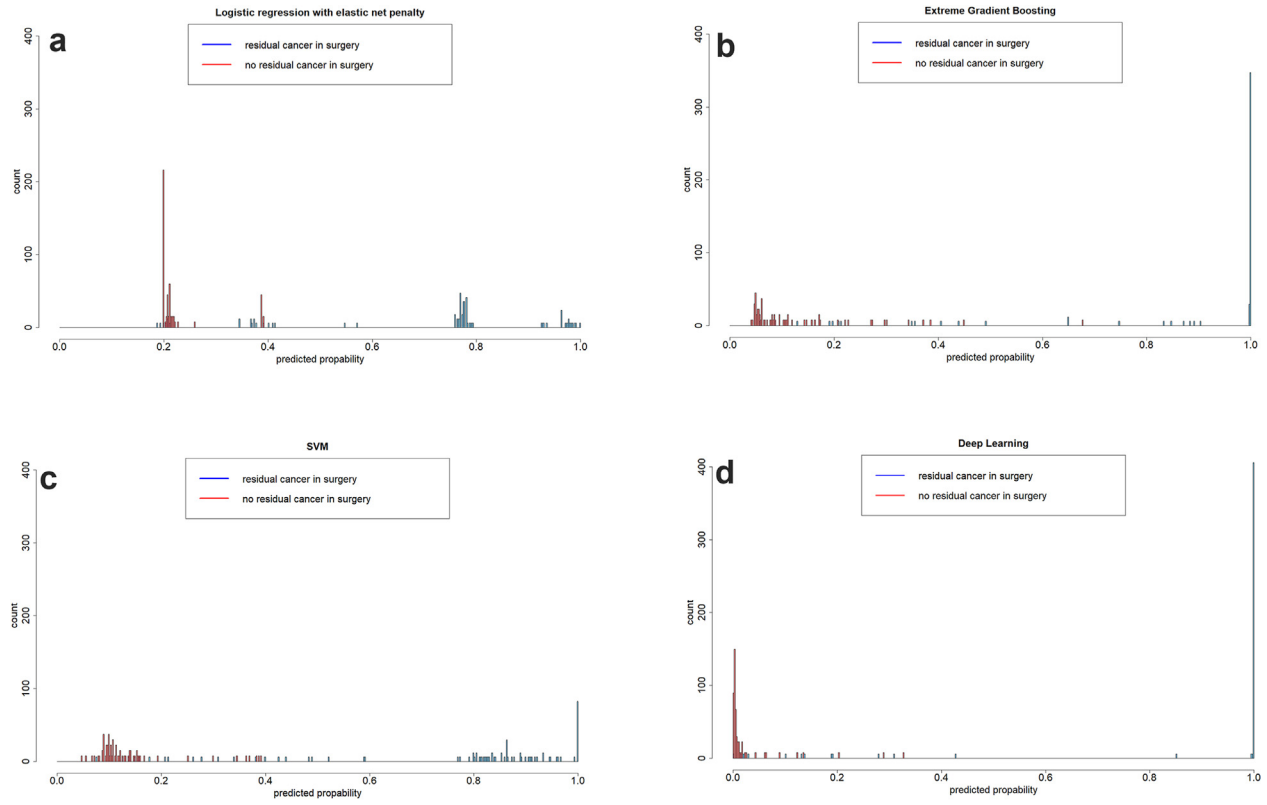

Fig. 3. Calibration of the four multivariate algorithms. Calculated probability for residual tumor after neoadjuvant treatment compared to actual residual cancer (blue) or no residual cancer (red) in surgery. (a) logistic regression with elastic net penalty. (b) Extreme Gradient Boosting tree. (c) Support-Vector-Machine. (d) deep neural network. . (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

algorithm predicts solely pCR of the primary tumor in the breast but not of the axillary nodes. The currently established diagnostic procedure in determining axillary disease for patients with early breast cancer is SLNB but trials are already underway to determine whether SLNB offers any benefit at all for clinically node negative patients [48]. Thus, our algorithm may be especially relevant and should be validated in cN0 patients with triple-negative or HER2-positive (maybe Luminal B) cancers, as they may be spared the operating room completely after adequate non-surgical staging of the breast and axilla. Besides the influence of tumor biology, our analysis provides additional insights which patients are at risk of having residual disease after NST. Following the SHAP and LIME value interpretation for the XGBoost tree model and the neural network, patients with larger lesion diameters on imaging before (and after) NST, older patients, patients with invasive lobular cancers, and patients with an accompanying *in situ* component (as assessed by the initial diagnostic biopsy) are at higher risk to have residual disease after NST. This is in line with previous research concluding that these patients generally respond worse to NST [49–51] and strengthens the credibility in local interpretation methods like SHAP [31] and LIME [38] which provide insights into the predictions made by complex black-box models.

Fourth, we are aware of the fact that different definitions of breast pCR exist (ypT0/ypTis). As we are discussing the possibility to omit breast surgery for women with breast pCR after NST, we chose the most conservative definition of breast pCR as absence of both invasive and *in situ* tumor cells.

Fifth, our choice to analyze handcrafted variables instead of an image recognition analysis of the respective radiologic and pathologic images (which has been the main application of complex algorithms like neural networks so far), was informed by the previous literature: Both imaging [10,11] and VAB [14–17] resulted in high false-negative rates (missed residual disease). Image recognition algorithms in other areas have yielded performance comparable to physician experts but usually do not surpass them [52]. Also in the prediction of breast pCR, image recognition analysis of MRI showed suboptimal accuracy [53]. The sole analysis of breast images or histopathologic VAB slides may not account for the complex and individual interaction of patient, tumor, tumor response, and

VAB characteristics. Although handcrafted variables may be subject to interrater variability (as every medical procedure as standardized as it might be), these variables followed standard clinical protocols as specified in Table 1. Combining the populations of four clinical trials caused some heterogeneity of variables (e.g., ultrasound and mammography and/or MRI were used as image modalities as applicable in clinical routine; differences in the distribution of baseline variables like age, tumor size, and receptor status) which may however better represent the clinical reality and strengthen the confidence in the generalizability of our findings.

5. Conclusion

A multivariate algorithm might accurately select breast cancer patients without residual disease after neoadjuvant treatment. This finding may pave the way to study omission of surgery in these patients in the future.

Authors' contributions

André Pfob: Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Roles/Writing – original draft; Writing – review & editing. Chris Sidey-Gibbons: Conceptualization; Formal analysis; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Visualization; Writing – review & editing. Han-Byoel Lee: Conceptualization; Data curation; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Writing – review & editing.Marios Konstantinos Tasoulis: Conceptualization; Data curation; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Writing – review & editing.Vivian Koelbel: Data curation; Software; Writing – review & editing.Michael Golatta: Conceptualization; Data curation; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Writing – review & editing.Gaiane M. Rauch: Conceptualization; Data curation; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Writing – review & editing.Benjamin D. Smith: Conceptualization; Data curation; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Writing – review & editing.Vicente Valero: Conceptualization; Data curation; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Writing – review & editing.Wonshik Han: Conceptualization; Data curation; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Writing – review & editing.Fiona MacNeill: Conceptualization; Data curation; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Writing – review & editing.Walter Paul Weber: Conceptualization; Methodology; Supervision; Writing – review & editing.Geraldine Rauch: Conceptualization; Funding acquisition; Methodology; Supervision; Validation; Writing – review & editing. Henry M. Kuerer: Conceptualization; Data curation; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Writing – review & editing.Joerg Heil: Conceptualization; Data curation; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Visualization; Roles/Writing – original draft; Writing – review & editing.

Data were analyzed by André Pfob and Chris Sidey-Gibbons.

The first draft of the manuscript was drafted by André Pfob and Joerg Heil. All authors contributed in collecting data and finalizing the manuscript, and all authors approved the final version.

Funding

There was no funding for this article. The single trials were supported by a Cancer Center Support Grant from the National Institutes of Health (NIH) (CA16672), a NIH P30 grant (CA016672), and funding form the MD Anderson Clinical Research Funding Award Program [13], the German German Research Foundation (DFG) (GZ:HE 6824/5-1) [14], the Seoul National University Hospital Research Fund (30-2016-0250), and the Institute for Information and Communications Technology Promotion (IITP) grant funded by the Korea government (2018-0-00861) [18].

Writing assistance

We thank Stephanie Deming of Scientific Publications, Research Medical Library, MD Anderson Cancer Center, for editing the manuscript.

Ethics committee approval

All trials and their combined analysis were approved by the respective institutional review boards and ethics committees. All human participants gave written informed consent. All trials were registered at clinicaltrials.gov with the identifiers NCT02455791, NCT02948764, NCT03273426, and NCT02575612.

Data sharing statement

Will individual participant data be available (including data dictionaries)?

Yes.

What data in particular will be shared?

Individual participant data that underlie the results reported in this article, after deidentification (text, tables, figures, and appendices).

What other documents will be available?

Study Protocol.

When will data be available (start and end dates)?

Immediately following publication. No end date.

With whom?

Researchers who provide a methodologically sound proposal.

For what types of analyses?

To achieve aims in the approved proposal.

By what mechanism will data be made available? Proposals should be directed to joerg.heil@med.uni-heidelberg.de.

To gain access, data requestors will need to sign a data access agreement.

Code availability

We used the open source "R" programming language (©The R Foundation for Statistical Computing). The machine learning frameworks used in this study (TensorFlow and Keras) are available at https://github.com/tensorflow/tensorflow and https://github.com/keras-team/keras.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We thank Stephanie Deming of Scientific Publications, Research Medical Library, MD Anderson Cancer Center, for editing the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejca.2020.11.006.

References

[1] Murphy BL, Day CN, Hoskin TL, Habermann EB, Boughey JC. Neoadjuvant chemotherapy use in breast cancer is greatest in excellent responders: triple-negative and HER2+ subtypes. Ann Surg Oncol 2018;25:2241–8. https://doi.org/10.1245/s10434-018-6531-5.

[2] van Ramshorst MS, van der Voort A, van Werkhoven ED, Mandjes IA, Kemper I, Dezentje VO, et al. Neoadjuvant chemotherapy with or without anthracyclines in the presence of dual HER2 blockade for HER2-positive breast cancer (TRAIN-2): a multicentre, open-label, randomised, phase 3 trial. Lancet Oncol 2018;19:1630–40. https://doi.org/10.1016/S1470-2045(18)30570-9.

[3] Santonja A, Sánchez-Muñoz A, Lluch A, Chica-Parrado MR, Albanell J, Chacón JI, et al. Triple negative breast cancer subtypes and pathologic complete response rate to neoadjuvant chemotherapy. Oncotarget 2018;9:26406–16. https://doi.org/10.18632/oncotarget.25413.

[4] Haque W, Verma V, Hatch S, Suzanne Klimberg V, Brian Butler E, Teh BS. Response rates and pathologic complete response by breast cancer molecular subtype following neoadjuvant chemotherapy. Breast Canc Res Treat 2018;170:559–67. https://doi.org/10.1007/s10549-018-4801-3.

[5] Goorts B, van Nijnatten TJ, de Munck L, Moossdorff M, Heuts EM, de Boer M, et al. Clinical tumor stage is the most important predictor of pathological complete response rate after neoadjuvant chemotherapy in breast cancer patients. Breast Canc Res Treat 2017;163:83–91. https://doi.org/10.1007/s10549-017-4155-2.

[6] Veronesi U, Cascinelli N, Mariani L, Greco M, Saccozzi R, Luini A, et al. Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. N Engl J Med 2002;347:1227–32. https://doi.org/10.1056/NEJMoa020989.

[7] Fisher B, Anderson S, Bryant J, Margolese RG, Deutsch M, Fisher ER, et al. Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. N Engl J Med 2002;347:1233–41. https://doi.org/10.1056/NEJMoa022152.

[8] Veronesi U, Paganelli G, Galimberti V, Viale G, Zurrida S, Bedoni M, et al. Sentinel-node biopsy to avoid axillary dissection in breast cancer with clinically negative lymph-nodes. Lancet 1996;349:1864–7. https://doi.org/10.1016/S0140-6736(97)01004-0.

[9] Giuliano AE, Ballman KV, McCall L, Beitsch PD, Brennan MB, Kelemen PR, et al. Effect of axillary dissection vs No axillary dissection on 10-year overall survival among women with invasive breast cancer and sentinel node metastasis: the ACOSOG Z0011 (alliance) randomized clinical trial. J Am Med Assoc 2017;318:918–26. https://doi.org/10.1001/jama.2017.11470.

[10] Fowler AM, Mankoff DA, Joe BN. Imaging neoadjuvant therapy response in breast cancer. Radiology 2017;285:358–75. https://doi.org/10.1148/radiol.2017170180.

[11] Rauch GM, Adrada BE, Kuerer HM, van la Parra RF, Leung JW, Yang WT. Multimodality imaging for evaluating response to neoadjuvant chemotherapy in breast cancer. AJR Am J Roentgenol 2017;208:290–9. https://doi.org/10.2214/ajr.16.17223.

[12] Heil J, Schaefgen B, Sinn P, Richter H, Harcos A, Gomez C, et al. Can a pathological complete response of breast cancer after neoadjuvant chemotherapy be diagnosed by minimal invasive biopsy? Eur J Canc 2016;69:142–50. https://doi.org/10.1016/j.ejca.2016.09.034.

[13] Kuerer HM, Rauch GM, Krishnamurthy S, Adrada BE, Caudle AS, DeSnyder SM, et al. A clinical feasibility trial for identification of exceptional responders in whom breast cancer surgery can Be eliminated following neoadjuvant systemic therapy. Ann Surg 2018;267:946–51. https://doi.org/10.1097/sla.0000000000002313.

[14] Heil J, Pfob A, Sinn H-P, Rauch G, Bach P, Thomas B, et al. Diagnosing pathologic complete response in the breast after neoadjuvant systemic treatment of breast cancer patients by

minimal invasive biopsy. Ann Surg 2020;272. https://doi.org/10.1097/SLA.0000000000004246.
[15] Tasoulis MK, Lee H-B, Yang W, Pope R, Krishnamurthy S, Kim S-Y, et al. Abstract GS5-04: accuracy of post-neoadjuvant chemotherapy image-guided breast biopsy to predict the presence of residual cancer: a multi-institutional pooled analysis. Canc Res 2020;80. https://doi.org/10.1158/1538-7445.sabcs19-gs5-04. GS5-04.
[16] Basik M, Cecchini RS, Santos JFDL, Umphrey HR, Julian TB, Mamounas EP, et al. Abstract GS5-05: primary analysis of NRG-BR005, a phase II trial assessing accuracy of tumor bed biopsies in predicting pathologic complete response (pCR) in patients with clinical/radiological complete response after neoadjuvant chemotherapy (NCT) to exp. Canc Res 2020;80. https://doi.org/10.1158/1538-7445.sabcs19-gs5-05. GS5-05-GS5-05.
[17] Vrancken Peeters M-JTFD, van Loevezijn A, van der Noordaa ME, van Duijnhoven FH, Loo CE, van Werkhoven E, et al. Abstract GS5-06: towards omitting breast surgery in patients with a pathologic complete response after neoadjuvant systemic treatment: interim analysis of the MICRA trial (Minimally Invasive Complete Response Assessment). Canc Res 2020;80. https://doi.org/10.1158/1538-7445.sabcs19-gs5-06. GS5-06-GS5-06.
[18] Lee HB, Han W, Kim SY, Cho N, Kim KE, Park JH, et al. Prediction of pathologic complete response using image-guided biopsy after neoadjuvant chemotherapy in breast cancer patients selected based on MRI findings: a prospective feasibility trial. Breast Canc Res Treat 2020;182:97–105. https://doi.org/10.1007/s10549-020-05678-3.
[19] Wöckel A, Festl J, Stüber T, Brust K, Stangl S, Heuschmann PU, et al. Interdisciplinary screening, diagnosis, therapy and follow-up of breast cancer. Guideline of the DGGG and the DKG (S3-level, AWMF registry number 032/045ol, december 2017) - Part 1 with recommendations for the screening, diagnosis and therapy of breast Ca. Geburtshilfe Frauenheilkd 2018;78:927–48. https://doi.org/10.1055/a-0646-4522.
[20] Schwartz LH, Litière S, de Vries E, Ford R, Gwyther S, Mandrekar S, et al. RECIST 1.1-Update and clarification: from the RECIST committee. Eur J Canc 2016;62:132–7. https://doi.org/10.1016/j.ejca.2016.03.081.
[21] Liu Y, Chen PHC, Krause J, Peng L. How to read articles that use machine learning: users' guides to the medical literature. JAMA, J Am Med Assoc 2020;322:1806–16. https://doi.org/10.1001/jama.2019.16489.
[22] Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open 2016;6:e012799. https://doi.org/10.1136/bmjopen-2016-012799.
[23] Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:55–63. https://doi.org/10.7326/M14-0697.
[24] Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol 2019;19:1–18. https://doi.org/10.1186/s12874-019-0681-4.
[25] Pfob A, Mehrara B, Nelson J, Wilkins EG, Pusic A, Sidey-Gibbons C. Towards data-driven decision-making for breast cancer patients undergoing mastectomy and reconstruction: prediction of individual patient-reported outcomes at two-year follow-up using machine learning. J Clin Oncol 2020;38. https://doi.org/10.1200/JCO.2020.38.15_suppl.520. 520–520.
[26] Sidey-Gibbons C, Asaad M, Pfob A, Boukovalas S, Lin Y-L, Offodile A. Machine learning algorithms to predict financial toxicity associated with breast cancer treatment. J Clin Oncol 2020;38. https://doi.org/10.1200/JCO.2020.38.15_suppl.2047. 2047–2047.
[27] Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med 1997;16:385–95. https://doi.org/10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3.
[28] Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Software 2010;33:1–22.
[29] Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat 2001;29:1189–232. https://doi.org/10.1214/aos/1013203451.
[30] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. 17- August-2016 Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. 2016;13:785–94. https://doi.org/10.1145/2939672.2939785. New York, NY, USA: Association for Computing Machinery.
[31] Lundberg S, Lee S-I. A unified approach to interpreting model predictions. http://arxiv.org/abs/1705.07874; 2017.
[32] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20:273–97. https://doi.org/10.1007/bf00994018.
[33] Vapnik V, Golowich S, Smola A. Support vector method for function approximation, regression estimation and signal processing. Adv Neural Inform Process Syst 1997;9:281–7.
[34] Lecun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44. https://doi.org/10.1038/nature14539.
[35] Hahnloser RHR, Sarpeshkar R, Mahowald MA, Douglas RJ, Seung HS. Digital selection and analogue amplification coexist in a cortex- inspired silicon circuit. Nature 2000;405:947–51. https://doi.org/10.1038/35016072.
[36] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. Proc. Fourteenth Int. Conf. Artif. Intell. Stat. PMLR 2011;15315–323.
[37] Dozat T. Incorporating Nesterov momentum into Adam. Proc. Fourth Int. Conf. Learn. Represent. 2016.
[38] Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": explaining the predictions of any classifier. https://arxiv.org/abs/1602.04938; 2016.
[39] Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170:51–8. https://doi.org/10.7326/M18-1376.
[40] Gärtner R, Jensen MB, Nielsen J, Ewertz M, Kroman N, Kehlet H. Prevalence of and factors associated with persistent pain following breast cancer surgery. J Am Med Assoc 2009;302:1985–92. https://doi.org/10.1001/jama.2009.1568.
[41] Flanagan MR, Zabor EC, Romanoff A, Fuzesi S, Stempel M, Mehrara BJ, et al. A comparison of patient-reported outcomes after breast-conserving surgery and mastectomy with implant breast reconstruction. Ann Surg Oncol 2019;26:3133–40. https://doi.org/10.1245/s10434-019-07548-9.
[42] Heil J, Kuerer HM, Pfob A, Rauch G, Sinn HP, Golatta M, et al. Eliminating the breast cancer surgery paradigm after neoadjuvant systemic therapy: current evidence and future challenges. Ann Oncol 2020;31:61–71. https://doi.org/10.1016/j.annonc.2019.10.012.
[43] Heil J, Pfob A, Kuerer HM. De-escalation towards omission is the tipping point of individualizing breast cancer surgery. Eur J Surg Oncol 2020;46:1543–5. https://doi.org/10.1016/j.ejso.2020.03.208.
[44] Masuda N, Lee SJ, Ohtani S, Im YH, Lee ES, Yokota I, et al. Adjuvant capecitabine for breast cancer after preoperative chemotherapy. N Engl J Med 2017;376:2147–59. https://doi.org/10.1056/NEJMoa1612645.
[45] von Minckwitz G, Huang CS, Mano MS, Loibl S, Mamounas EP, Untch M, et al. Trastuzumab emtansine for residual invasive HER2-positive breast cancer. N Engl J Med 2019;380(7):617–28. https://doi.org/10.1056/NEJMoa1814017.
[46] Heil J, Pfob A. Patients should be the tipping point of individualizing breast cancer surgery: commentary on 'Eliminating the breast cancer surgery paradigm after neoadjuvant systemic therapy: current evidence and future challenges. Ann Oncol 2020;31:1264. https://doi.org/10.1016/j.annonc.2020.05.021.

[47] Caballero C, Piccart M. Important considerations prior to elimination of breast cancer surgery after neoadjuvant systemic therapy: listening to what our patients want. Ann Oncol 2020; 31:1083–4. https://doi.org/10.1016/j.annonc.2020.04.009.

[48] Reimer T. Omission of SLNB in triple-negative and HER2-positive breast cancer patients with rCR and pCR in the breast after NAST. Clin Identifier NCT04101851 [n.d].

[49] van la Parra RF, Kuerer HM. Selective elimination of breast cancer surgery in exceptional responders: historical perspective and current trials. Breast Cancer Res 2016;18:28. https://doi.org/10.1186/s13058-016-0684-6.

[50] Sun S, van la Parra RFD, Rauch GM, Checka C, Tadros AB, Lucci A, et al. Patient selection for clinical trials eliminating surgery for HER2-positive breast cancer treated with neoadjuvant systemic therapy. Ann Surg Oncol 2019;26:3071–9. https://doi.org/10.1245/s10434-019-07533-2.

[51] Von Minckwitz G, Darb-Esfahani S, Loibl S, Huober J, Tesch H, Solbach C, et al. Responsiveness of Adjacent ductal carcinoma in situ and changes in HER2 status after neoadjuvant chemotherapy/trastuzumab treatment in early breast cancer-results from the GeparQuattro study (GBG 40). Breast Canc Res Treat 2012;132:863–70. https://doi.org/10.1007/s10549-011-1621-0.

[52] Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Heal 2019;1:e271–97. https://doi.org/10.1016/s2589-7500(19)30123-2.

[53] Liu Z, Li Z, Qu J, Zhang R, Zhou X, Li L, et al. Radiomics of multiparametric MRI for pretreatment prediction of pathologic complete response to neoadjuvant chemotherapy in breast cancer: a multicenter study. Clin Canc Res 2019;25:3538–47. https://doi.org/10.1158/1078-0432.CCR-18-3190.

[54] Prabhu JS, Korlimarla A, Desai K, Alexander A, Raghavan R, Anupama CE, et al. A majority of low (1-10%) er positive breast cancers behave like hormone receptor negative tumors. J Canc 2014;5:156–65. https://doi.org/10.7150/jca.7668.