Journal of Pathology Informatics 18 (2025) 100448

Contents lists available at ScienceDirect

Journal of Pathology Informatics

journal homepage: www.elsevier.com/locate/jpi

Deep learning predicts the effect of neoadjuvant chemotherapy for patients with triple negative breast cancer

B. Sturm [a,b], P. Lock [c], D. Kumar [d], W.A.M. Blokx [e], J.A.W.M. van der Laak [a,*]

[a] Department of Pathology, Radboud University Medical Center, Nijmegen, the Netherlands
[b] Pathan, Rotterdam, the Netherlands
[c] Kentalis, Goes, the Netherlands
[d] Aiforia Technologies Plc, Helsinki, Finland
[e] Department of Pathology, Division of Laboratories, Pharmacy and Biomedical Genetics, University Medical Center Utrecht, Utrecht, the Netherlands

ARTICLE INFO

Keywords:
Deep learning
Artificial intelligence
WSI
Breast cancer
Chemotherapy
Prediction

ABSTRACT

Background: Triple negative breast cancer (TNBC) is an aggressive subcategory of breast cancer with poor prognosis and high risk of recurrence after treatment. In a subset of cases systemic chemotherapy is offered before surgery, so called neoadjuvant chemotherapy (NAC), to downstage the disease resulting in 40–50% of cases to a pathological complete response. Meanwhile, patients receiving NAC suffer from toxic side effects and in a proportion of patients a significant amount of residual tumor remains. This study aims to predict the outcome of NAC with deep learning technology based on the microscopic morphological characteristics in whole slide images of hematoxylin and eosin (H&E) slides from the pre-operative tumor biopsy before chemotherapy.

Methods: A convolutional neural network was trained on 221 H&E-stained biopsies of carcinoma of no special type from 205 patients scanned at 40×. Cases were divided in three cohorts, with a good, moderate, or bad response to NAC based on the EUSOMA scoring according to the pathology report of the subsequent tumor surgery specimen. We defined good, moderate, and bad response as residual tumor <10%, 10–50%, and 50%, respectively. Manual segmentation of the tumor area was performed comprising invasive carcinoma with a small rim of surrounding benign tissue. The model was tested on 52 new biopsies of 50 patients. Because of the relative low number of moderate and bad responder cases, and to achieve a better discrimination for potential visual biomarkers, the moderate and bad response cohorts were merged.

Results: The predictive performance of the model was calculated by means of the area under the receiver operator curve (AUC ROC). 95% Confidence intervals (CIs) were calculated for better understanding of the range of values. In the test set, the AUC ROC performance score was 0.696 with a CI of 0.532–0.861.

Conclusion: This proof-of-concept study shows that H&E pre-operative biopsies from TNBC, by means of deep learning technology, contain valuable information having predictive value for the outcome of NAC resulting in an AUC value of 0.696 outperforming a predictive AUC value of 0.63 based on structured clinical data of histological tumor grade, TILs, and ki-67 known from the literature.

Introduction

The incidence of breast cancer continues to be high and morbidity and mortality from the disease remains a significant health problem. Choice of breast cancer treatment is based on biopsy findings in combination with clinical information (tumor size and lymph node status based on clinical imaging). A pathologist subtypes and grades the tumor based on the hematoxylin and eosin (H&E) tissue slide of the biopsy. Therapeutic biomarkers are routinely performed: estrogen receptor (ER), progesterone receptor (PR), and HER2 expression, using ancillary techniques. Based on the clinical information in combination with the pathological biopsy findings the best-suited treatment is offered to the patient. In a subset of cases, systemic chemotherapy is offered before surgery, so called neoadjuvant chemotherapy (NAC). In primary breast cancer, the therapeutic objective of NAC is to shrink the tumor and downstage the disease. In advanced disease, NAC may convert inoperable to operable disease. Currently, the role of NAC has expanded to include patients with early-stage breast cancer that allows to perform a wide local tumor excision instead of a mastectomy, with a higher rate of clear margins, improving cosmetic outcomes and reducing post-operative complications such as lymphedema.[1,2] Meanwhile, data have shown that the response to chemotherapy is a strong predictive marker for the risk of recurrence and therefore a surrogate endpoint of

* Corresponding author at: Dept. of Pathology 824, Radboud University Medical Center, P.O.Box 9101, 6500HB Nijmegen, the Netherlands.
E-mail address: jeroen.vanderlaak@radboudumc.nl (J.A.W.M. van der Laak).

http://dx.doi.org/10.1016/j.jpi.2025.100448
Received 14 July 2024; Received in revised form 17 April 2025; Accepted 8 May 2025
Available online 14 May 2025

favorable survival, especially in triple negative breast carcinoma.[3] Triple negative breast cancer (TNBC) is an aggressive subcategory of breast cancer, comprising 10–15% of cases, with poor prognosis and high risk of recurrence after treatment.[4–6] After administering NAC, 40–50% of TNBC cases have no residual tumor left in the resection specimen, so called pathological complete response (pCR).[2,6,7] Patients receiving NAC, in general for the duration of 6 months from date of biopsy,[8] often suffer from toxic effects which is likely to worsen their prognosis and the therapy comes with financial costs. Therefore, predicting good vs poor response to NAC can reduce unnecessary harm in patients by sparing them ineffective and potentially toxic NAC treatment, and by performing tumor resection without the delay of around 4–6 months due to chemotherapy.[3,9]

Diverse biomarkers were proposed to predict NAC response in TNBC: tumor size, histological grade, tumor proliferation score based on Ki-67 staining and tumor-infiltrating lymphocytes (TILs). In general, small tumor size, high grade, high Ki-67 expression, and high TILs are positively related to pCR in NAC setting.[3,6] In recent years, genetic biomarker tests have become available that predict a response to NAC (e.g., Mammaprint and Oncotype Dx). However, these test have substantial drawbacks in time investment, high cost, and are not applicable for TNBC.[10]

Increasingly, pathology departments digitally scan all tissue slides.[11] Availability of these digital images enables the use of deep learning technologies, which were shown to be useful in for example breast cancer detection and grading in digitalized H&E tissue slides.[12–15] Histological images harbor information on tumor characteristics reflecting underlying molecular processes and biological disease mechanisms.[16] Humans have been able to correlate visual data from H&E tissue slides with disease behavior in breast cancer resulting in the Nottingham Bloom and Richardson grading system that nowadays guides treatment of breast cancer patients.[17] The visual data in tissue slides are abundant and complex, and likely also contains information that is not (yet) visible to the human eye. Artificial intelligence may be able to extract visual data that correlates with molecular processes and disease progression.

In this proof-of-concept study, a convolutional neural network (CNN) was trained and tested to predict the NAC response based on the morphological information in the H&E slide of the tumor biopsy specimen before NAC and surgery in TNBC of no special type (NST). The CNN was trained in a naive way/weakly supervised manner without addressing already known biomarkers. This method allows a prediction from the digital image in seconds compared to laborious grading by the pathologist and (possibly in the future) financially costly genetic predictive testing taking several days or weeks.

Methods and materials

Selection

All cases from the pathology archive of Pathan B.V. were collected concerning formalin-fixed paraffin-embedded H&E-stained slides of biopsies of triple negative carcinoma NST of the breast that received NAC for the duration of $\geq$4.5 months before surgery from the years 2017 until 2022. In this study, the exact administered medicine (combination) and number of cycles for NAC was not known. Therefore, a surrogate complete period of NAC was defined as at least 6 months between the date of the (last acquired) pre-operative biopsy and the date of surgery. In the case of moderate and bad responders, a NAC cut-off period of at least 5.5 and 4.5 months, respectively, was set as a viable period in which a good response was not to be expected and were included. Existing H&E tissue sections from cases from the years 2017 until 2020 derived from five hospitals were H&E restained because a high number of the earlier years showed faded staining. In cases without a precursor ductal carcinoma in situ (DCIS) lesion, a GATA3 staining was performed to exclude tumor metastatic disease from outside the breast. Carcinoma NST with metaplastic features were excluded as well. As a separate test set, 52 biopsy cases of 50 patients from the year 2022 were collected according to the criteria described above. All cases were scanned with a Philips ultra fast scanner (UFS; serial number 228 (Philips, Amsterdam, The Netherlands)) at 40$\times$ (pixel size 0.25 μm) and uploaded to the Aiforia's commercially available cloud-based platform (Aiforia Technologies Plc, Helsinki, Finland).

Training and testing cohorts

221 Cases from 205 patients were divided in three cohorts of good, moderate, and bad responders to NAC based on the EUSOMA scoring according to the pathology report of the tumor surgery specimen. We defined good response as residual tumor of less than 10%, moderate response as 10–50% residual tumor, and bad response as 50% residual tumor.

The level from the biopsy specimen with the greatest amount of tumor tissue within the slide was selected for training of a CNN. Within the Aiforia cloud-based platform, the tumor area was manually delineated by a pathologist (BS). Precise invasive tumor regions were annotated, including a small rim of surrounding area and excluding surrounding whitespace background, benign tissue, DCIS, artifacts (e.g., folded tissue, scratches, squeezed areas, and out of focus areas), extensive fibrosis/sclerosis, and necrosis. Next, for CNN training, all regions in a slide were associated with a label, expressing the specific patient cohort (good, moderate, or bad response to NAC). The CNN was trained using patches of 200 $\times$ 200 μm. Due to the relative low number of moderate and bad responder cases, and to achieve a better discrimination for potential visual biomarkers, the moderate and bad response groups were merged for training. Further details on CNN training can be found in the supplementary information table. In Fig. 1, an overview of the study design is provided.

A 5-fold cross-validation (CV) was performed on the cases from the years 2017–2021. The CV groups were slightly modified to correct for multiple cases of the same patient and for proportions of good, moderate, and bad response to NAC, to prevent bias. For testing of the model, 52 biopsy cases of 50 patients from the year 2022 were collected which were not present in the training data.

Evaluation and statistical analysis

The Aiforia software outputs, for every manually drawn region of interest, the surface areas classified, respectively, as good responder or moderate and bad responder. To calculate a patient level likelihood (ranging from 0 to 1) for being a moderate and bad responder, we calculated the area fraction classified as moderate and bad responder over the sum of all annotated regions per patient. To evaluate the predictive performance of the model, this likelihood was used in the area under the receiver operator curve (AUC ROC) analysis. To identify possible histological features that may predict NAC response, we analyzed areas with high AI confidence (90%).

Results

Table 1 gives an overview of the global performance of the AI model regarding the training and validation of the CV groups and the test set. The predictive AUC ROC performance score regarding the CV groups ranged from 0.541 to 0.730 (mean 0.616) and in the test set was 0.696, presented in Fig. 2 and Table 2. An example of a result in the WSI of the algorithm is illustrated in Fig. 3 featuring the precise manual segmentation of the invasive tumor area. To give an indication of the clinical usefulness of the model developed in this study, applying this model, we can identify 23.1% of moderate and bad responders at the cost of falsely classifying 2.6% of the good responders as moderate and bad responders. Based on high AI predicted confidence, it seems that areas with syncytial cell groups with at most minor cribriform tubule formation, mainly bi-phasic/2 populations of epithelial cells and presence of leukocytes (lymphocytes plasma cells neutrophils) and sometimes nuclear debris and/or apoptotic cells may be predictive features for good response to NAC. Regarding histological features that seem to be predictive for moderate or bad response to NAC relatively abundant stroma (sclerotic/desmoplastic) with some degree of small spindle cell nuclei if low cellularity stroma, white areas (i.e., cleft artifacts, tubules, or small vessels), compact/blue round epithelial areas and areas

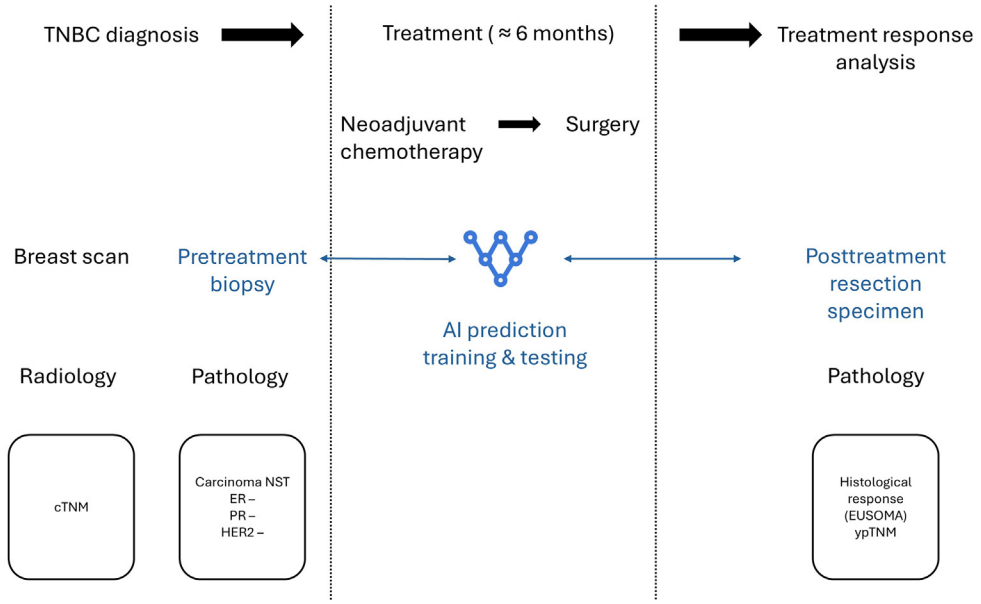

Fig. 1. Overview of the study design.

with lower leukocyte density compared to comparable tumor areas with higher leukocyte density. Examples of these morphological features are provided in Fig. 4.

Discussion

This study shows the feasibility of using AI to extract valuable information from digitalized H&E tissue slides that correlates with NAC response. This finding is concordant with other recent studies. Li et al.[3] developed an AI-based pCR-score using data from 540 patients, that was found to outperform basic biomarkers of sTILs and tumor subtype. This score, which was not specifically developed for TNBC, obtained an AUC of 0.847 on an independent validation set. The study design of Li et al. is different from the present study with respect to inclusion of multiple tumor subtypes (HR+, HR–, HER2+, and HER2–), patients without metastasis and the application of automated tumor segmentation that may have an impact on the results. Naylor et al.[18] trained a CNN on a dataset of 336 patients using the residual cancer burden (RCB) based on the pathology report of the surgical resection specimen as the outcome of NAC. Two prediction settings were investigated: (1) pCR (no residuum) vs RCB (some residuum) and (2) pCR-RCB-I vs RCB-II-III, that is supposed to be clinically more relevant, as it is informative of a patient's prognosis. The 5-fold nested CV achieved a mean AUC ROC performance of 0.654 ± 0.049 and 0.601 ± 0.049, respectively, as the best performing model in the prediction of response to NAC for TNBC. Our study is comparable in outcome classification design (<10% vs 10–100% residual tumor according to EUSOMA criteria) to predict patient's prognosis of pCR – RCB-I vs RCB-II - RCB-III achieving a comparable mean AUC ROC performance of 0.616. Aswolinskiy et al.[19] showed, concerning TNBC cases a promising result on computational TILs and inflamed tumor ratio with an AUC performance of 0.77 and 0.71, respectively, the results from an internal validation set (n = 66; n pCR = 29), although the number of cases is small. Unfortunately, the so-called PROACTING biomarkers from WSI H&E slides show no single biomarker having absolute statistical significance to be sufficient for patient care to benefit from. Our study differs with respect to Aswolinskiy et al. approach not focusing on already known biomarkers but training the CNN in a naive/agnostic way, possibly including known biomarkers as well as unknown (combination of) biomarkers. Ogier du Terrail et al.[20] in a proof-of-concept study based on 650 patients showed that machine learning models relying on WSI can predict pCR to NAC therapy in TNBC. The study showed an AUC ROC performance of 0.66 on average when applying the best federated learning method outperforming the clinical prediction with an AUC ROC performance of 0.63, based on structured clinical data of histological tumor grade, TILs, and ki-67. The available data also show that with the

Table 1
Overview of the global performance of the AI model. CV = cross-validation group, FP = False positive, FN = False negative. FN and FP is calculated per class. The total area error is calculated for all the classes combined and explains the relative high error values.

Training, validation & test groups	Total area error (%)	Area error (mm^2)	Area total (mm^2)	False positive (%)	Area error FP (mm^2)	Area total FP (mm^2)	False negative (%)	Area error FN (mm^2)	Area total FN (mm^2)	Number of slides
Training CV2, 3, 4 & 5	1.79	34.06	1899.46	0.44	16.78	3798.92	0.45	17.28	3798.92	177
Training CV1, 3, 4 & 5	81.27	1568.68	1930.12	20.30	783.74	3860.24	20.33	784.94	3860.24	175
Training CV1, 2, 4 & 5	1.15	23.17	2010.22	0.28	11.33	4020.44	0.29	11.85	4020.44	177
Training CV1, 2, 3 & 5	1.02	19.83	1943.88	0.25	9.67	3887.75	0.26	10.16	3887.75	177
Training CV1, 2, 3 & 4	0.75	13.86	1845.87	0.18	6.70	3691.75	0.19	7.16	3691.75	178
Training CV1, 2, 3, 4 & 5 (all cases 2017–2021)	1.63	38.90	2390.40	0.40	19.17	4780.80	0.41	19.73	4780.80	221
Validation CV1 based on Training CV2, 3, 4 & 5	78.55	399.86	509.03	19.62	199.74	1018.06	19.66	200.12	1018.06	44
Validation CV2 based on Training CV1, 3, 4 & 5	91.50	435.69	476.17	22.86	217.70	952.33	22.89	217.99	952.33	46
Validation CV3 based on Training CV1, 2, 4 & 5	92.12	366.89	398.27	23.01	183.27	796.53	23.05	183.62	796.53	44
Validation CV4 based on Training CV1, 2, 3 & 5	72.88	336.99	462.41	18.20	168.30	924.83	18.24	168.70	924.83	44
Validation CV5 based on Training CV1, 2, 3 & 4	73.03	410.86	562.61	18.46	207.67	1125.22	18.06	203.19	1125.22	43
Validation Test 22 based on Training CV1, 2, 3, 4 & 5	70.46	334.87	475.27	17.57	167.00	950.55	17.66	167.87	950.55	52

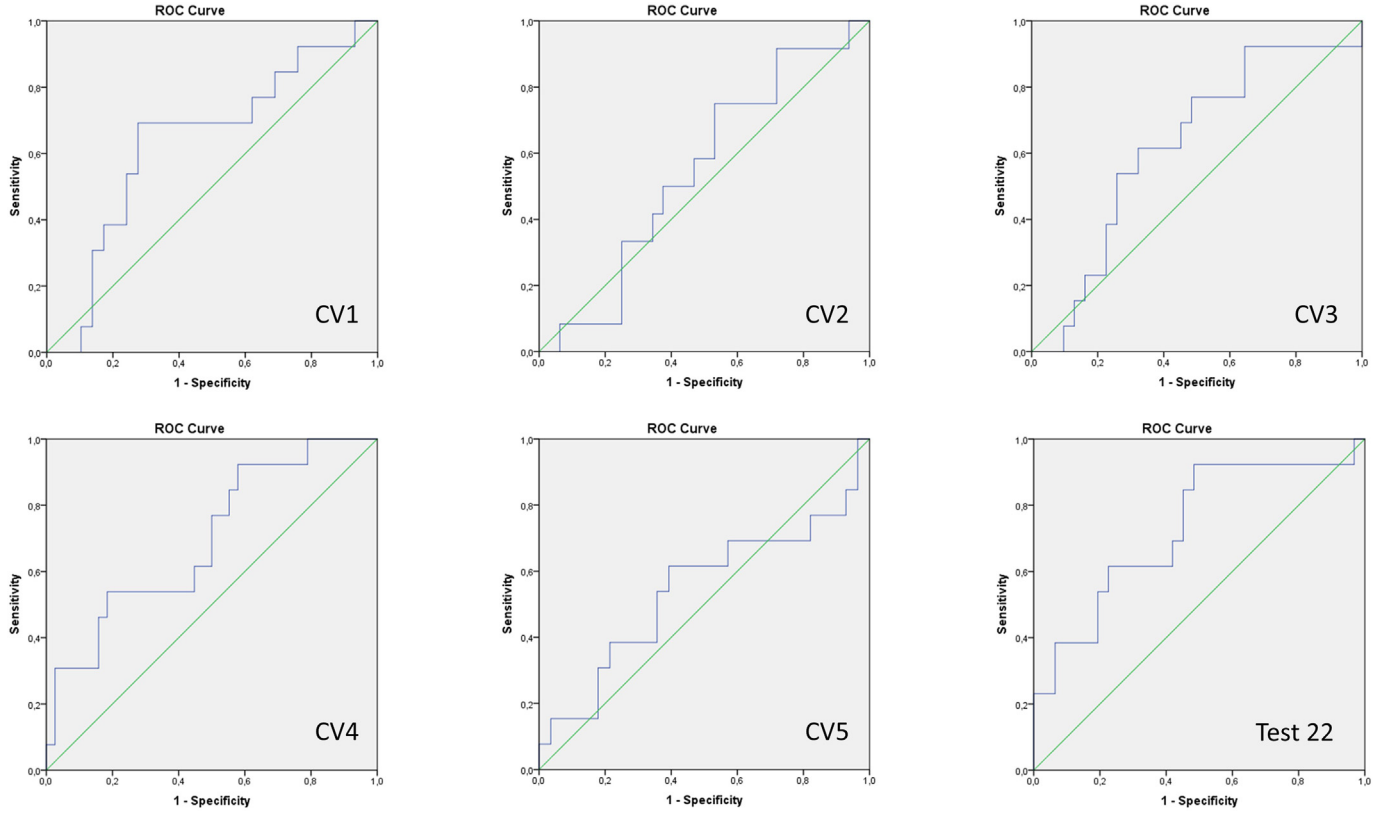

Fig. 2. ROC curves. CV = cross-validation group.

approach of a bimodal score distribution on all tiles that tiles associated with high and low scores are presenting clear histological patterns. By means of extensive expert annotations of predefined criteria, it was possible to highlight specific patterns associated with response or nonresponse in the entire test sets. The presence of TILs and hemorrhage along with necrosis are associated with overall high scores and probability to response, whereas the presence of apocrine tumor cells, fibrosis, and non-cohesive tumor cells are associated with low scores and high risk of poor NAC response. The authors state that most of those patterns are already well known in the community (i.e., TILs, apocrine or non-cohesive tumor cells, or hemorrhage) and therefore, confirm the biological relevance of their algorithm. Some other patterns, like the presence of fibrosis already suggested by the literature, require a deeper analysis to understand the link made by their algorithm between this pattern and a low probability of response. The above-mentioned features may have contributed in our study design by segmenting the invasive tumor area with a small rim of surrounding benign tissue including TIL-s, necrosis, fibrosis, and hemorrhage in combination with a relative high level of detail of 200 μm^2 field of view. However, not the aim of our study, it would be interesting to investigate if unknown biomarkers are present in the data.

Regarding the above-mentioned studies, we need to stress that these studies focus on pCR, whereas our study focuses on good response defined as less than 10% residual tumor that includes pCR cases. The rational for this is that tissue sampling always has a chance viable tumor is not sampled because as a rule of thumb in a resection specimen a pathologist for practical reasons samples 1 section per cm tumor diameter and in some cases, it is not possible to find the exact region the tumor was supposed to be in the resection specimen resulting in possible sampling bias. Subsequently, AUC values in this study compared to other studies are not exactly comparable, except for the study of Naylor et al. Already mentioned above regarding study design, it appeared that in a 5-fold CV setting by Naylor et al., the pCR prediction of NAC response gives on average probably a slightly higher ROC AUC value than the pCR – RCB-I vs RCB-II - RCB-III prediction when analyzing fig. 8 in their scientific article, showing 14 different algorithm approaches. However, exact values are not presented for each algorithm implying comparing remains currently an educated guess.

In our study, to compensate for the relatively small number of cases while trying to achieve a high yield of discriminatory visual biomarkers, precise manual invasive tumor area segmentation was done excluding areas with carcinoma in situ (CIS), extensive necrosis, extensive fibrosis, and tissue artifacts (e.g., tissue folds, out of focus areas, squeezed areas, dust particles among others). The rational to exclude big areas of fibrosis/sclerosis and necrosis is the potential pitfall for the algorithm to classify this as viable tumor but in fact is a secondary change in the tumor possibly introducing bias/confounding factor, although some fibrosis and necrosis was segmented in the small rim surrounding invasive tumor cells. In the EUSOMA classification, residual CIS without residual infiltrative tumor after NAC is regarded as a pCR. Therefore, CIS was excluded because it is a pre-neoplastic lesion and in a subset of cases does not respond to NAC. Also, some TNBCs have a DCIS-like cytomorphology and growth pattern that complicate the CNN training even further introducing a confounding factor when CIS is included in tumor segmentation. Metaplastic carcinoma and other breast carcinoma of special type, that can be included in the group of TNBC, are rare and diverse in cytomorphology ranging in the case of metaplastic carcinoma from squamous cell carcinoma to different

Table 2
Number of biopsies per group and cohort. AUC with 95% CI per group.

Validation & test groups	Good response	Moderate response	Bad response	AUC	95% CI
CV1	32	6	6	0.547	0.366–0.728
CV2	33	6	7	0.623	0.445–0.801
CV3	31	6	7	0.541	0.334–0.748
CV4	31	6	7	0.730	0.559–0.900
CV5	30	6	7	0.637	0.452–0.822
Test 22	39	9	4	0.696	0.532–0.861

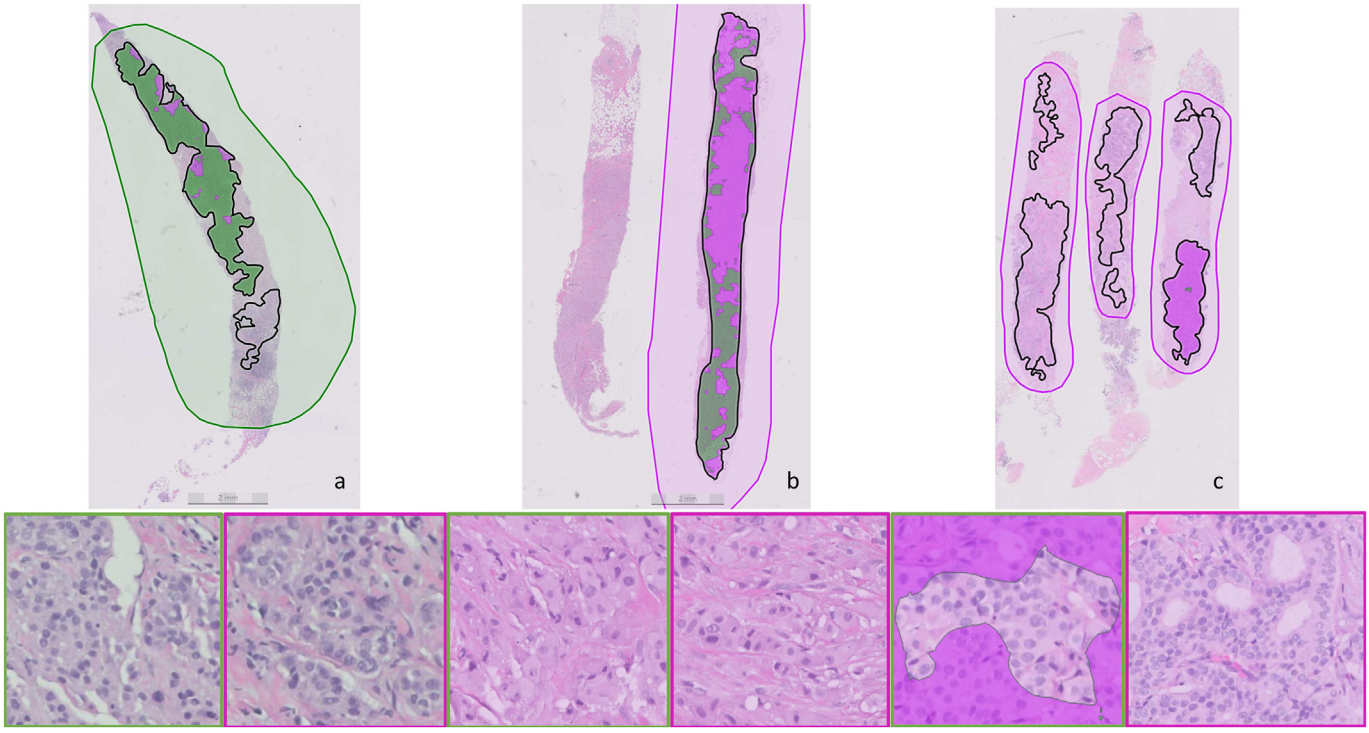

Fig. 3. Example of AI prediction in a case of good (a), moderate (b), and bad (c) response to NAC. Areas or line contour in green represent predicted good and in purple moderate or bad response, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

types of sarcoma, and for that matter were excluded from this study, although a subset does respond well to NAC.[21,22]

Because therapeutic protocols change over time only cases from 2017 and onwards were collected in this study because from 2017 and onwards, a new NAC protocol was introduced in The Netherlands. A limitation of the present study is the fact that the exact administered medicine (combination) and number of NAC cycles was not known. The defined completion period of NAC in this study is based on standard clinical NAC protocols, with treatment typically spanning approximately 6 months from the date of biopsy.[8] In our dataset, 19 cases fell in the range of 4.5–6 months

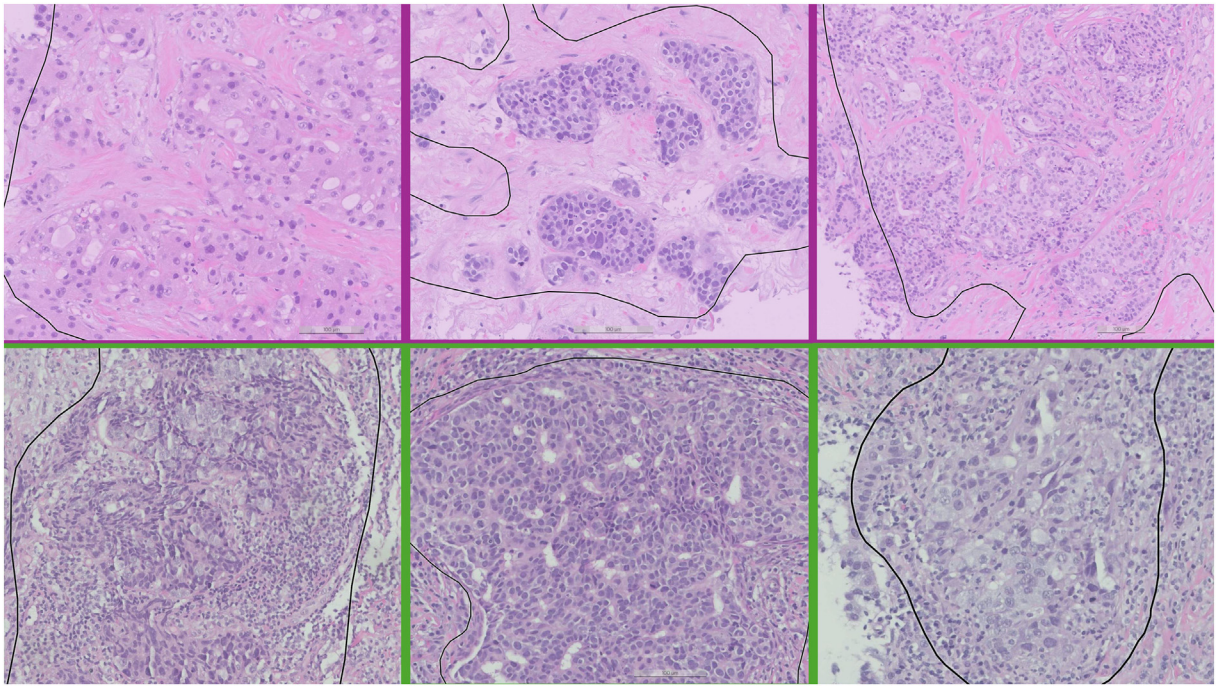

Fig. 4. Examples detailed morphology high AI confidence (90%) prediction of moderate or bad response (row above; line contour purple) and good response (row below; line contour green). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

measured from date of biopsy to date of surgery. 16 cases involving 15 patients were excluded upfront from the primary set of cases due to an abbreviated NAC period (<4.5 months). We believe this exclusion is methodologically justified, as shorter treatment durations may introduce confounding variables, thereby compromising the reliability of the study. A higher AUC value may be achieved when training is done on cases with more precise clinical knowledge of the tumor size, lymph node status, NAC regimen, and number of cycles in a specific patient.

For the last decades, histological tumor grading (tubule formation, nuclear pleomorphism, and mitotic count) has been a simple and inexpensive method having important prognostic value for breast cancer patients and has a prominent role in guiding treatment. The prognostic significance has been widely studied, showing among other things, that histological grade in conjunction with disease state (consisting of tumor size and lymph node status) could improve outcome conditions. It has been shown that the prognostic value of histological grade was independent of tumor size and lymph node status and that grade is complimentary and equivalent in impact magnitude to lymph node status, which is widely regarded as a major prognostic factor in breast cancer. Moreover, the independent prognostic effect of histological grading on survival (as well as lymph node status and tumor size) is long-lasting (10 years). Regarding the prognostic contribution of the three constituents of grade, several studies have shown that the mitotic count is the most important variable, followed by nuclear atypia and then tubule formation.[6] AI has the potential to outperform the contemporary clinical use of histological tumor grade, ki-67, TILs, and gene expression profiling in conjunction with tumor size and lymph node status. This will eventually alleviate the workload of pathologists and decrease the cost of healthcare.

We found that some morphological features, foremost TILs, and sclerotic stroma, seem to be prognostic for NAC response and are in line with results from other studies.[20,23,24] Ogier du Terrail et al. found hemorrhage, TILs, and necrosis as predictive of pCR. Apocrine change, fibrosis, and dyscohesive tumor cells being predictive of bad response. Hacking et al. quantified the stromal and tumor features in a WSI-based multi-omic (WSI, clinical, and pathological) machine learning model and found that high collagenous stroma was best associated with lower pCR rates. Fisher et al. provided evidence that the inter-relationships between TILs, stroma, adipocytes, and tumor cells can predict NAC response in patients with TNBC.

In our study, similar tumor morphology from different patients showed a similar prediction from the algorithm although a different real-world therapeutic outcome. This underlines the fact that tumor cytomorphology alone may have only relative value in therapeutic outcome in relation to many other intrinsic patient factors, i.e., biological metabolism factors, immunological factors, physical (e.g., body mass index), and mental condition among others.[25–29] We acknowledge that our study lacks a detailed analysis of model performance regarding misclassified cases and failure modes as it was designed as a proof-of-concept study. Therefore, future studies need to incorporate so-called big data analysis for treatment response prediction, including in-patient factors, which may play a guiding role in this area in the future, whereas laborious tumor grading, scoring TILs, and expensive gene expressing profiling may become obsolete.

Finally, an AI-driven diagnostic tumor WSI analysis may serve as a tool for therapeutic decision-making, guiding the selection of NAC or adjuvant chemotherapy based on predictive modeling of treatment outcomes. Specifically, a favorable prognostic assessment could support the choice of NAC followed by surgical tumor resection, whereas an unfavorable prediction may warrant surgical tumor resection first followed by consideration of adjuvant chemotherapy as an alternative strategy.

Conclusion

This proof-of-concept study shows that AI can extract valuable information from H&E-stained pre-operative biopsies, having predictive value for the outcome of NAC in TNBC. Future studies need to prove if a higher AUC value may be achieved when precise clinical information is included in CNN training.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Bart Sturm reports financial support was provided by Infodatek Group. Jeroen van der Laak reports a relationship with Aiosyn B.V. that includes: board membership and equity or stocks. Jeroen van der Laak was a member of the advisory boards of Philips, the Netherlands and ContextVision, Sweden, and received research funding from Philips, the Netherlands, ContextVision, Sweden, and Sectra, Sweden in the last 5 years. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpi.2025.100448.

References

1. Pinder SE, Provenzano E, Earl H, Ellis IO. Laboratory handling and histology reporting of breast specimens from patients who have received neoadjuvant chemotherapy. Histopathology 2007;50(4):409–417.
2. Asaoka M, Gandhi S, Ishikawa T, Takabe K. Neoadjuvant chemotherapy for breast cancer: past, present, and future. Breast Cancer (Auckl) 2020;14.1178223420980377.
3. Li F, Yang Y, Wei Y, et al. Deep learning-based predictive biomarker of pathological complete response to neoadjuvant chemotherapy from histological images in breast cancer. J Transl Med 2021;19(1):348.
4. Society AC, ed. Breast Cancer Facts & Figures 2022–2024; 2022. Atlanta.
5. d'Alessio V. Breast cancer: an aggressive variant triggers a hunt for cures. Horizon, the EU Research and Innovation Magazine. 2022. [Internet] https://projects.research-and-innovation.ec.europa.eu/en/horizon-magazine/breast-cancer-aggressive-variant-triggers-hunt-cures.
6. van den Ende NS, Nguyen AH, Jager A, Kok M, Debets R, van Deurzen CHM. Triple-negative breast cancer and predictive markers of response to neoadjuvant chemotherapy: a systematic review. Int J Mol Sci 2023;24(3).
7. Vaidya JS, Massarut S, Vaidya HJ, et al. Rethinking neoadjuvant chemotherapy for breast cancer. BMJ 2018;360, j5913.
8. Borstkanker. Triple negatieve borstkanker - Chemotherapie [Internet]. Available from: https://richtlijnendatabase.nl/gerelateerde_documenten/f/22750/Borstkanker%20-%20Triple%20negatieve%20borstkanker.pdf.
9. Zhao Y, Schaafsma E, Cheng C. Gene signature-based prediction of triple-negative breast cancer patient response to neoadjuvant chemotherapy. Cancer Med 2020;9(17):6281–6295.
10. Xin L, Liu YH, Martin TA, Jiang WG. The era of multigene panels comes? The clinical utility of Oncotype DX and MammaPrint. World J Oncol 2017;8(2):34–40.
11. Schwen LO, Kiehl TR, Carvalho R, Zerbe N, Homeyer A. Digitization of pathology labs: a review of lessons learned. Lab Investig 2023;103(11), 100244.
12. Gecer B, Aksoy S, Mercan E, Shapiro LG, Weaver DL, Elmore JG. Detection and classification of cancer in whole slide breast histopathology images using deep convolutional networks. Pattern Recogn 2018;84:345–356.
13. Mercan E, Mehta S, Bartlett J, Shapiro LG, Weaver DL, Elmore JG. Assessment of machine learning of breast pathology structures for automated differentiation of breast cancer and high-risk proliferative lesions. JAMA Netw Open 2019;2(8), e198777.
14. Sandbank J, Bataillon G, Nudelman A, et al. Validation and real-world clinical application of an artificial intelligence algorithm for breast cancer detection in biopsies. NPJ Breast Cancer 2022;8(1):129.
15. Wetstein SC, de Jong VMT, Stathonikos N, et al. Deep learning-based breast cancer grading and survival analysis on whole-slide histopathology images. Sci Rep 2022;12(1), 15102.
16. Mandair D, Reis-Filho JS, Ashworth A. Biological insights and novel biomarker discovery through deep learning approaches in breast cancer histopathology. NPJ Breast Cancer 2023;9(1):21.
17. van Dooijeweert C, van Diest PJ, Ellis IO. Grading of invasive breast carcinoma: the way forward. Virchows Arch 2022;480(1):33–43.
18. Naylor P, Lazard T, Bataillon G, et al. Prediction of treatment response in triple negative breast cancer from whole slide images. Front Signal Process 2022:2.
19. Aswolinskiy W, Munari E, Horlings HM, et al. PROACTING: predicting pathological complete response to neoadjuvant chemotherapy in breast cancer from routine

diagnostic histopathology biopsies with deep learning. Breast Cancer Res 2023;25 (1):142.
20. Ogier du Terrail J, Leopold A, Joly C, et al. Federated learning for predicting histological response to neoadjuvant chemotherapy in triple-negative breast cancer. Nat Med 2023;29(1):135–146.
21. Wong W, Brogi E, Reis-Filho JS, et al. Poor response to neoadjuvant chemotherapy in metaplastic breast carcinoma. NPJ Breast Cancer 2021;7(1):96.
22. Haque W, Verma V, Schwartz MR, et al. Neoadjuvant chemotherapy for metaplastic breast cancer: response rates, management, and outcomes. Clin Breast Cancer 2022;22 (5):e691–e699.
23. Hacking SM, Karam J, Singh K, et al. Whole slide image features predict pathologic complete response and poor clinical outcomes in triple-negative breast cancer. Pathol Res Pract 2023;246, 154476.
24. Fisher TB, Saini G, Rekha TS, et al. Digital image analysis and machine learning-assisted prediction of neoadjuvant chemotherapy response in triple-negative breast cancer. Breast Cancer Res 2024;26(1):12.
25. Lafourcade A, His M, Baglietto L, Boutron-Ruault MC, Dossus L, Rondeau V. Factors associated with breast cancer recurrences or mortality and dynamic prediction of death using history of cancer recurrences: the French E3N cohort. BMC Cancer 2018;18(1):171.
26. Arce-Salinas C, Aguilar-Ponce JL, Villarreal-Garza C, et al. Overweight and obesity as poor prognostic factors in locally advanced breast cancer patients. Breast Cancer Res Treat 2014;146(1):183–188.
27. Weiss SE, Tartter PI, Ahmed S, et al. Ethnic differences in risk and prognostic factors for breast cancer. Cancer 1995;76(2):268–274.
28. Wang X, Wang N, Zhong L, et al. Prognostic value of depression and anxiety on breast cancer recurrence and mortality: a systematic review and meta-analysis of 282,203 patients. Mol Psychiatry 2020;25(12):3186–3197.
29. de Souza BF, de Moraes JA, Inocenti A, dos Santos MA, Silva AE, Miasso AI. Women with breast cancer taking chemotherapy: depression symptoms and treatment adherence. Rev Lat Am Enfermagem 2014;22(5):866–873.