Deep Learning-Based Segmentation of Locally Advanced Breast Cancer on MRI in Relation to Residual Cancer Burden: A Multi-Institutional Cohort Study

Author name(s) withheld EMAIL(S) WITHHELD
Address withheld

Editors: Under Review for MIDL 2023

Abstract

While several methods have been proposed for automated assessment of breast-cancer response to neoadjuvant chemotherapy on breast MRI, limited information is available about their performance across multiple institutions. In this paper, we assess the value and robustness of nnU-Net-derived volumes of locally advanced breast cancer (LABC) on MRI to infer the presence of residual disease after neoadjuvant chemotherapy. An nnU-Net was trained to segment LABC on a single-institution training set and validated on a multi-center independent testing cohort. Based on resulting tumor volumes, an extremely randomized tree model was trained to assess residual cancer burden (RCB)-0/I vs. RCB-II/III. An independent model was developed using functional tumor volume (FTV). Models were tested on an independent testing cohort, response assessment performance and robustness across multiple institutions were assessed. Results show that nnU-Net accurately estimate changes in tumor load on DCE-MRI, that these changes associated with RCB after NAC, and that they are robust against variations between institutions.

Keywords: List of keywords, comma separated.

1. Introduction

Neoadjuvant chemotherapy (NAC) is increasingly used to treat patients with breast cancer as it allows monitoring of treatment response with the tumor *in situ* thus offers opportunity for more personalized treatment. The most sensitive modality to visualize tumor extent in three dimensions is dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI).

Methods for response monitoring on MRI range from manual assessment by radiologists to methods being investigated for fully automated analysis. Manual assessment has been shown to be predictive of pathological complete response (pCR), depending on tumor subtype (Janssen et al., 2022). Combinations of manually selecting regions of interest (ROI) and semi-automatic thresholding have also been proposed, including a semi-automated method to establish functional tumor volume (FTV) (Hylton et al., 2016). Fully automated data-driven methods to assess response to NAC on MRI have also been proposed, using radiomics or deep learning (Choi et al., 2019; Comes et al., 2021; Joo et al., 2021). Little is known about the robustness of these methods across multiple institutions.

This study aimed to establish whether nnU-Net accurately assesses changes in tumor load on DCE-MRI that are associated with residual cancer burden (RCB) after NAC robust to variations between institutions and MRI scanners. Secondly, whether such model is in agreement with the relationship between FTV and RCB.

© 2023 CC-BY 4.0, Anonymous.

2. Methods

2.1. Datasets

The training cohort consisted of 105 consecutively included breast cancer cases treated with neoadjuvant chemotherapy (NAC) in a single institution. Patients underwent two MRI examinations: The first at baseline prior to NAC, and the second either midway through the chemotherapy schedule, or immediately before the second-to-last cycle of chemotherapy. The second MRI examination was defined as the follow-up examination. Either 1.5 T or 3 T Philips scanners were used in this set. The independent testing cohort consisted of 54 consecutively included breast cancer cases treated with NAC in four institutions. The baseline scan before any treatment were used as well as the follow-up scans after all cycles of NAC but prior to surgery. Imaging in the testing cohort was performed exclusively on 3 T scanners, both Philips and Siemens. All examinations had a pre-contrast scan and up to five post-contrast scans acquired at intervals of 55 to 89 seconds, fat-supressed was used. Ground-truth annotations of all scans were derived from a previously reported histopathology-validated semi-automated region grower (Alderliesten et al., 2007).

To evaluate the response to NAC on histopathology, the residual cancer burden RCB was derived from the final post-surgery resection specimens following the methodology described by Symmans *et al* (Symmans et al., 2007). The RCB score was dichotomized into two categories: RCB categories RCB-0 (i.e. pCR, pathological complete response) and RCB-I were defined as good responders to NAC. Conversely, categories RCB-II and RCB-III were considered bad responders.

2.2. Response assessment

The ground truth segmentations were used to train a 3D nnU-Net CNN (Isensee et al., 2021). Input to the nnU-Net were the precontrast DCE series and five postcontrast DCE MRI (i.e. six channels in total). To evaluate segmentation performance, two-fold cross-validation was performed on the training set, while for final evaluation of response assessment performance, the network was trained on the entire training set.

To compare nnU-Net to a previously validated method for response assessment, Functional Tumor Volume (FTV) was also calculated per breast from each MRI examination following the description by Newitt *et al.* (Newitt et al., 2014).

A model based on extremely randomized trees was fit to assess tumor response to NAC (Geurts et al., 2006). Three input candidate predictors were used: the lesion volume determined on the follow-up scan, tumor subtype (either HER2+, HER2-/ER+ or triple negative) and the difference in tumor volume between baseline and follow-up scan. The end point was tumor response expressed as the dichotomized RCB. The area under the receiver operator curve (AUC) was used as measure of model performance. Five-fold nested cross-validation was performed for hyperparameter tuning and internal model validation. Two separate models were trained, one where the volumes were determined using the previously trained nnU-Net, the second one using FTV.

Table 1: Performance of tumor response assessment in terms of residual cancer burden (RCB) on a per-hospital basis for the nnU-Net and FTV segmentation methods. Numbers presented are areas under the receiver operator curve (AUC).

	nnU-Net	**FTV**
Hospital 1 ($n = 12$)	0.63	0.71
Hospital 2 ($n = 21$)	0.74	0.75
Hospital 3 ($n = 19$)	0.79	0.81
Hospital 4 ($n = 2$)	1.00	1.00

3. Results and conclusion

The median (interquartile range (IQR)) cross-validated Dice score from the nnU-Net, was 0.87 (0.62-0.93). Pearson's correlation coefficient between volumes derived from the nnU-Net and the ground truth was R=0.95 (fold 1: $R = 0.93$, fold 2: $R = 0.97$). The correlation between the nnU-Net-derived volume and FTV in the training cohort was $R = 0.74$ for the baseline scan, $R = 0.72$ for the follow-up scan, and $R = 0.80$ for all scans combined. All correlations were statistically significant ($P < 0.05$). In the testing cohort, the median (IQR) AUC of the response assessment model was 0.76 (0.71-0.84) for nnU-Net-derived tumor volumes and 0.77 (0.74-0.86) for FTV. There was no significant difference in AUC between the two models ($p = 0.66$). Per hospital performance varied, with the worst performance associated with the hospital from the training set (Hospital 1) (Table 1).

We conclude that nnU-Net can accurately estimate changes in tumor load on DCE-MRI and that these changes are associated with RCB after NAC. The response assessment is on par with that derived using FTV, a previously validated method, but it is fully automated and therefore observer independent. The performance of the model appears to be robust to variations in scan parameters across multiple institutions, proving the versatility of the method.

Acknowledgments

Acknowledgments withheld.

References

Tanja Alderliesten, Angelique Schlief, Johannes Peterse, Claudette Loo, Hendrik Teertstra, Sara Muller, and Kenneth Gilhuijs. Validation of semiautomatic measurement of the extent of breast tumors using contrast-enhanced magnetic resonance imaging. *Investigative Radiology*, 42:42–49, 2007. ISSN 00209996. doi: 10.1097/01.rli.0000248849.99014.7e.

Woo Jung Choi, Hak Hee Kim, Joo Hee Cha, Hee Jung Shin, and Eun Young Chae. Comparison of pathologic response evaluation systems after neoadjuvant chemotherapy in breast cancers: Correlation with computer-aided diagnosis of mri features. *American Journal of Roentgenology*, 213:944–952, 2019. ISSN 15463141. doi: 10.2214/AJR.18.21016. URL http://www.ncbi.nlm.nih.gov/pubmed/31237439.

Maria Colomba Comes, Annarita Fanizzi, Samantha Bove, Vittorio Didonna, Sergio Diotaiuti, Daniele La Forgia, Agnese Latorre, Eugenio Martinelli, Arianna Mencattini, Annalisa Nardone, Angelo Virgilio Paradiso, Cosmo Maurizio Ressa, Pasquale Tamborra, Vito Lorusso, and Raffaella Massafra. Early prediction of neoadjuvant chemotherapy response by exploiting a transfer learning approach on breast dce-mris. *Scientific Reports*, 11:1–12, 2021. ISSN 20452322. doi: 10.1038/s41598-021-93592-z. URL https://doi.org/10.1038/s41598-021-93592-z.

Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely randomized trees. *Machine Learning*, 63:3–42, 4 2006. ISSN 0885-6125. doi: 10.1007/s10994-006-6226-1. URL http://www.ncbi.nlm.nih.gov/pubmed/31076546http://link.springer.com/10.1007/s10994-006-6226-1.

Nola M. Hylton, Constantine A. Gatsonis, Mark A. Rosen, Constance D. Lehman, David C. Newitt, Savannah C. Partridge, Wanda K. Bernreuter, Etta D. Pisano, Elizabeth A. Morris, Paul T. Weatherall, Sandra M. Polin, Gillian M. Newstead, Helga S. Marques, Laura J. Esserman, and Mitchell D. Schnall. Neoadjuvant chemotherapy for breast cancer: Functional tumor volume by mr imaging predicts recurrencefree survival-results from the acrin 6657/calgb 150007 i-spy 1 trial. *Radiology*, 279:44–55, 12 2016. ISSN 15271315. doi: 10.1148/radiol.2015150013. URL https://pubs.rsna.org/doi/abs/10.1148/radiol.2015150013.

Fabian Isensee, Paul F. Jaeger, Simon A.A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18:203–211, 2021. ISSN 15487105. doi: 10.1038/s41592-020-01008-z. URL http://dx.doi.org/10.1038/s41592-020-01008-z.

L. M. Janssen, B. M. den Dekker, K. G. A. Gilhuijs, P. J. van Diest, E. van der Wall, and S. G. Elias. Mri to assess response after neoadjuvant chemotherapy in breast cancer subtypes: a systematic review and meta-analysis. *npj Breast Cancer*, 8:107, 9 2022. ISSN 2374-4677. doi: 10.1038/s41523-022-00475-1. URL https://www.nature.com/articles/s41523-022-00475-1.

Sunghoon Joo, Eun Sook Ko, Soonhwan Kwon, Eunjoo Jeon, Hyungsik Jung, Ji-Yeon Kim, Myung Jin Chung, and Young-Hyuck Im. Multimodal deep learning models for the prediction of pathologic response to neoadjuvant chemotherapy in breast cancer. *Scientific reports*, 11:18800, 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-98408-8. URL https://doi.org/10.1038/s41598-021-98408-8http://www.ncbi.nlm.nih.gov/pubmed/34552163http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC8458289.

David C. Newitt, Sheye O. Aliu, Neil Witcomb, Gal Sela, John Kornak, Laura Esserman, and Nola M. Hylton. Real-time measurement of functional tumor volume by mri to assess treatment response in breast cancer neoadjuvant clinical trials: Validation of the aegis ser software platform. *Translational Oncology*, 7:94–100, 2014. ISSN 19365233. doi: 10.1593/tlo.13877.

W. Fraser Symmans, Florentia Peintinger, Christos Hatzis, Radhika Rajan, Henry Kuerer, Vicente Valero, Lina Assad, Anna Poniecka, Bryan Hennessy, Marjorie Green, Aman U. Buzdar, S. Eva Singletary, Gabriel N. Hortobagyi, and Lajos Pusztai. Measurement of residual breast cancer burden to predict survival after neoadjuvant chemotherapy. *Journal of Clinical Oncology*, 25:4414–4422, 2007. ISSN 0732183X. doi: 10.1200/JCO.2007.10.6823.