Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240606

Deep Learning Based Automatic Fibroglandular Tissue Segmentation in Breast Magnetic Resonance Imaging Screening

Guelsuem PEHLIVAN[a], Carl Mathis WILD[b,c], Julia BAUMGARTL[d], Dennis HARTMANN[a], Nina DITSCH[b], Frank KRAMER[a] and Dominik MUELLER [a,c,1]
[a]*IT-Infrastructure for Translational Medical Research, University of Augsburg, Germany*
[b]*Department of Gynecology and Obstetrics, University Hospital Augsburg, Germany*
[c]*Department of Radiology, University Hospital Augsburg, Germany*
[d]*Institute for Digital Medicine, University Hospital Augsburg, Germany*

Abstract. In light of the global increase in breast cancer cases and the crucial importance of the density of fibroglandular tissue (FGT) in assessing risk and predicting the course of the disease, the accurate measurement of FGT emerges as a significant challenge in diagnostic imaging. The current study focuses on the automatic segmentation of breast glandular tissue in MRI scans using a deep learning model. The aim is to establish a solid foundation for the development of methods for the precise quantification of fibroglandular tissue. For this purpose, the publicly available 'Duke Breast Cancer MRI' dataset was systematically processed to train a deep neural network model utilizing the nnU-Net ('no-new-Net') framework, which was then subjected to a quantitative evaluation. The results show the following macro-averaged metrics with standard deviation: Dice Similarity Coefficient 0.827 ± 0.152, accuracy 0.997 ± 0.003, sensitivity 0.825 ± 0.158, and specificity 0.999 ± 0.001. The effectiveness of our model in segmenting FGT is underscored by the high values of the Dice coefficient, Accuracy, Sensitivity, and Specificity, which reflect the precision and reliability of our results. The findings of this study lay a solid foundation for developing automated methods to quantify FGT. Our research efforts, especially driven by clinical studies at the University Hospital Augsburg, are focused on further exploring and validating these potentials.

Keywords. Deep Learning; Gynecology; Magnetic Resonance Imaging; Fibroglandular Tissue; Image Segmentation

1. Introduction

Breast cancer is one of the most common cancers worldwide and exhibits a high mortality rate [1-3]. Remaining fibroglandular tissue after skin-sparing mastectomy increases the risk of new carcinoma development or recurrence [1-3]. Accurate quantification of FGT plays a crucial role in diagnosis and treatment planning [1-3].

---

[1] Corresponding author: Dominik Müller, IT Infrastructure for Translational Medical Research, Alter Postweg 101, 86159 Augsburg, Germany; E-mail: dominik.mueller@informatik.uni-augsburg.de.

However, current methods for measuring the amount of FGT, especially post-operative interventions, are often inadequate and subjective. Therefore, there is an urgent need for improved objective approaches to FGT determination. Recent advancements in artificial intelligence and machine deep learning suggest that the use of more sophisticated and objective techniques can significantly enhance medical diagnostics and therapy, while also creating reliable quantification procedures. Against this backdrop, the present study aims to employ deep learning for automated segmentation of FGT in Magnetic Resonance Imaging (MRI) images. This is intended to establish reliable quantification procedures that are crucial for making informed postoperative treatment decisions in breast cancer patients.

2. Methods

In this section, we explain the preparation process of fibroglandular tissue data for the deep learning model, which was used for the automatic segmentation of breast glandular tissue. This tissue, also referred to as 'fibroglandular tissue', included the non-fatty tissue crucial for breast cancer diagnostics. The focus was on the importance of high-quality datasets for machine learning, which were essential for the training, performance, and evaluation of models.

2.1. Dataset Description

The ‘Duke Breast Cancer MRI Dataset’ was a collection of dynamic contrast-enhanced magnetic resonance images (DCE-MRI) from 922 patients with invasive breast cancer, collected between 2000 and 2014 at the Duke University [4]. These images included preoperative MRI scans performed with either 1.5 T or 3 T scanners in the prone position [4, 5]. The dataset contained non-fat-suppressed T1-weighted sequences, fat-suppressed gradient-echo T1-weighted pre-contrast sequences, and typically three to four post-contrast sequences [4, 5]. The annotation process was carried out by radiologists using bounding boxes to mark tumor locations on the DCE-MRI images [4]. For 271 patients, the annotations were performed by a panel of six fellowship-trained radiologists, with each radiologist randomly assigned to a study [4]. For the remaining 651 patients, the largest biopsied lesion was annotated by a panel of four fellowship-trained radiologists, modifying the procedure as necessary to access the PACS system [4]. To image data and breast glandular tissue segmentations, the dataset included extensive demographic, clinical, pathological, genomic, and treatment information, as well as technical MRI details [4]. The breast glandular tissues comprised 356 MRI slices.

2.2. Deep Learning Model

The nnU-Net framework was developed to enable efficient processing and assignment of segmentation masks to corresponding MRI scans [6]. The Duke dataset was divided into two segments, with 285 scans along with their segmentation masks used for training and the remaining 71 scans for evaluation. The training of nnU-Net was conducted using a five-fold cross-validation, dividing the training set into five equal subgroups to improve generalizability and more efficient data usage [4]. The framework archived relevant configurations, enhancing the reproducibility, analysis, further development, and automation of the model [6]. Specific hyperparameters were established, including a batch size of 14 and a patch size of 448 x 512 pixels, to ensure efficient utilization of GPU memory and stability during the training process [6]. The selection of these parameters depended on the requirements of the training dataset and hardware capacity [6]. During training, resampling adjusted image sizes to a uniform scale, aiding the model's learning by standardizing images, thus improving pattern recognition across datasets [6]. The loss function was defined by the combination of cross-entropy and the Dice Similarity Coefficient [6]. This blend aimed to enhance segmentation accuracy by accounting for both segment agreement and class-specific errors [6].

In evaluating segmentation algorithms in image analysis, we adhered to standardized performance evaluation guidelines [7, 8]. These metrics – Dice Similarity Coefficient, accuracy, sensitivity, and specificity – enabled us to assess the effectiveness and consistency of the algorithms by demonstrating how well the algorithm results aligned with actual data and what variability these results exhibited [7, 8].

3. Results

In this section, the performance of the trained deep learning model is presented. The macro-averaged scores for the evaluation criteria, along with their standard deviations, were as follows: Dice Similarity Coefficient0.827 ± 0.152, accuracy 0.997 ± 0.003, sensitivity 0.825 ± 0.158, and specificity 0.999 ± 0.001. These findings confirm the model's high precision and reliability in segmenting fibroglandular tissue within MRI scans. The strong overlap between the predicted and actual segmentations, as indicated by the Dice Coefficient, showcases the model's efficacy. Additionally, the high accuracy illustrates the nnU-Net model's exceptional precision in segmentation tasks. The sensitivity and specificity metrics emphasize the model's capability to accurately identify true positives and true negatives, respectively.

For a clearer visualization of the precision of results, an MRI breast scan is shown in Figure 1. This is followed by representations that showed the breast with the Ground Truth and the MRI scan with the algorithmically generated segmentation mask, respectively. In the lower right corner of Figure 1, a direct comparison of the two segmentation masks was made, where agreements were highlighted in green and discrepancies in red. Furthermore, the values for Sensitivity = 0.827 and Specificity = 0.997 were disclosed.

4. Discussion

An average Dice Similarity Coefficient of 0.827 indicates a pronounced agreement between the segmentation masks automatically generated by our model and the ground truth masks. This confirms the model's precise identification and delineation of breast glandular tissue in MRI scans. The Accuracy of 0.997 is in concordance with the high Dice Similarity Coefficient, reflecting the model's precision in differentiating breast glandular tissue from other tissue types. This demonstrates the model's robustness in producing segmentation results that are closely aligned with the ground truth, supporting its reliability for clinical use. A sensitivity of 0.825 ensures the correct identification of a significant portion of the existing breast glandular tissue. Moreover, a specificity of

0.999 effectively eliminates the non-presence of breast glandular tissue, contributing to minimizing the number of false-positive results and preventing both overdiagnoses and unnecessary invasive follow-up examinations.

The results impressively demonstrate that a deep neural network model based on the nnU-Net framework can efficiently differentiate between fibroglandular tissue (FGT) and other structures, as illustrated in Figure 1. This ability to distinguish is especially important for medical applications, as an accurate quantification of FGT significantly influences treatment strategies [1].

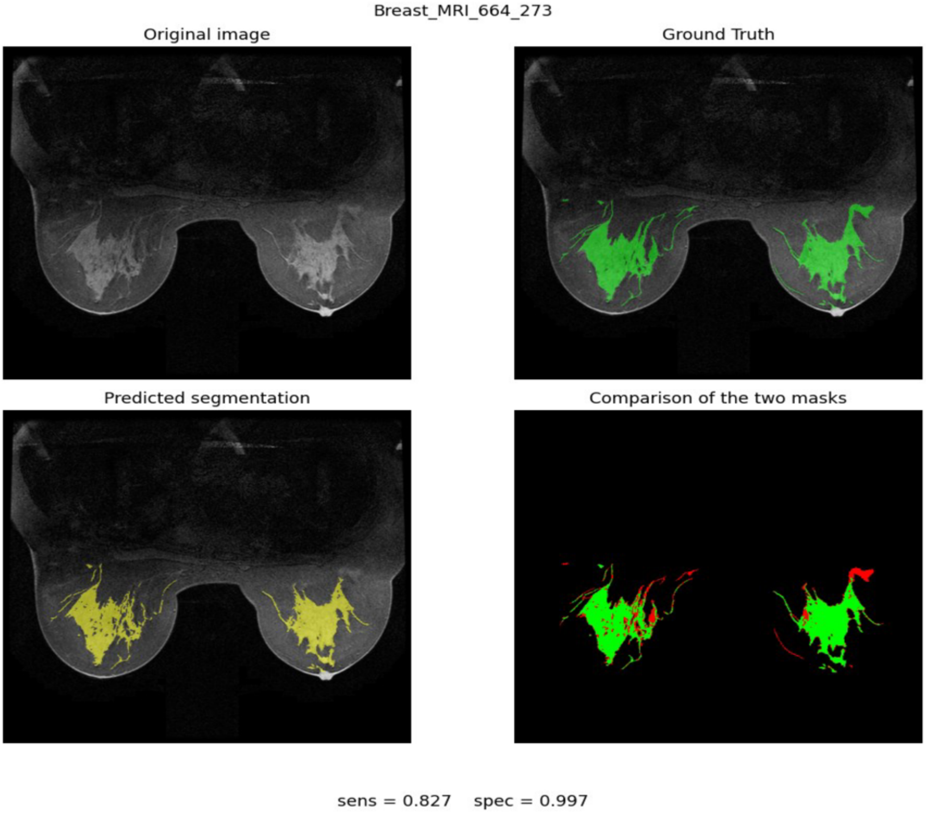

Figure 1. The illustration displays an MRI scan of the chest in four different presentations. The original scan is shown top left. Top right, the Ground Truth mask is presented in green, indicating the exact areas of interest. Bottom left shows the predicted segmentation in yellow. Bottom right compares both masks: green for correct predictions and red for discrepancies, highlighting over- or under-segmented areas. Beneath the image, the values for sens = sensitivity and spec = specificity are provided.

5. Conclusions

This study highlights the significant role of automated deep learning methods, particularly the nnU-Net framework, in the segmentation of fibroglandular tissue (FGT) in MRI scans. This technology, which enables effective differentiation between blood vessels, adipose tissue, and FGT, proves to be a promising tool for medical image analysis and precision medicine. The results achieved strong prediction capabilities that confirmed the ability of our model to perform precise and reliable segmentations. The insights gained from this study lay a solid foundation for future developments, which are to be particularly advanced by the University Hospital Augsburg. Nevertheless, we identified further potential by utilizing 3D models instead of our applied 2.5D strategy as recent literature suggests increased performance and robustness.

References

[1] Deutschmann C, Singer CF, Gschwantler-Kaulich D, Pfeiler G, Leser C, Baltzer PAT, et al. Residual fibroglandular breast tissue after mastectomy is associated with an increased risk of a local recurrence or a new primary breast cancer. BMC Cancer. 2023;23:281. doi: 10.1186/s12885-023-10764-y.

[2] Müller-Franzes- G, Müller-Franzes F, Huck L, Raaff V, Kemmer E, Khader F, et al. Fibroglandular tissue segmentation in breast MRI using vision transformers: a multi-institutional evaluation. Sci Rep. 2023;13(1):14207. doi: 10.1038/s41598-023-41331-x.

[3] Ponsiglione AM, Angelone F, Amato F, Sansone M. A statistical approach to assess the robustness of radiomics features in the discrimination of mammographic lesions. J Pers Med. 2023;13(7). doi: 10.3390/jpm13071104.

[4] Saha A, Harowicz MR, Grimm LJ, Kim CE, Ghate SV, Walsh R, Mazurowski MA. A machine learning approach to radiogenomics of breast cancer: a study of 922 subjects and 529 DCE-MRI features. Br J Cancer. 2018;119(4):508-516. doi:10.1038/s41416-018-0185-8.

[5] Buchanan CR, Muñoz Maniega S, Valdés Hernández MC, Ballerini L, Barclay G, Taylor AM, et al. Comparison of structural MRI brain measures between 1.5 and 3 T: Data from the Lothian Birth Cohort 1936. Hum Brain Mapp. 2021;42(12):3905-21. doi: 10.1002/hbm.25473.

[6] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods. 2021;18(2):203-211. doi:10.1038/s41592-020-01008-z.

[7] Maier-Hein L, Reinke A, Godau P, Tizabi MD, Buettner F, Christodoulou E, et al. Metrics reloaded: recommendations for image analysis validation. Nat Methods. 2024;21(1):195–212. doi:10.1038/s41592-023-02151-z.

[8] Müller D, Soto-Rey I, Kramer F. Towards a guideline for evaluation metrics in medical image segmentation. BMC Res Notes. 2022;15:210. doi:10.1186/s13104-022-06096-y.