Radiomics for the Discrimination of Infiltrative vs In Situ Breast Cancer

Benedetta Tafuri\textsuperscript{1,2}, Luana Conte*\textsuperscript{1,2}, Maurizio Portaluri\textsuperscript{3}, Alessandro Galiano\textsuperscript{4}, Eleonora Maggiulli\textsuperscript{5} and Giorgio De Nunzio\textsuperscript{1,2}

\textsuperscript{1}Laboratory of Biomedical Physics and Environment, Department of Mathematics and Physics "Ennio De Giorgi", University of Salento, Lecce, Italy
\textsuperscript{2}Laboratory of Interdisciplinary Research Applied to Medicine (DReAM), University of Salento and ASL (Local Health Authority), Lecce, Italy
\textsuperscript{3}ASL (Local Health Authority), Brindisi, and 'Di Summa-Perrino' Hospital, Operative Unit of Radiotherapy, Brindisi, Italy
\textsuperscript{4}ASL (Local Health Authority), Brindisi, and 'Di Summa-Perrino' Hospital, Operative Unit of Radiodiagnostics, Brindisi, Italy
\textsuperscript{5}ASL (Local Health Authority), Brindisi, and 'Di Summa-Perrino' Hospital, Operative Unit of Medical Physics, Brindisi, Italy

*Corresponding author: Luana Conte, Laboratory of Biomedical Physics and Environment, Laboratory of Interdisciplinary Research Applied to Medicine, Department of Mathematics and Physics, DReAM, P.O. "V.FAZZI", Pza F. Muratore n.1, Pal. Direz. Amministrativa, I piano, 73100 Lecce, Italy. Email: luana.conte@unisalento.it

Abstract

Breast cancer is the most common malignant tumor in women worldwide. Its early diagnosis relies on radiology and clinical evaluation, supplemented by biopsy confirmation. Technological advances in medical imaging, especially in the field of artificial intelligence, allow to address clinical challenges in cancer detection and classification, as well as in the assessment of treatment response, and in monitoring disease progression. Radiomics allows to extract features from images, related to tumor size, shape, intensity, and texture, providing comprehensive tumor characterization. In this paper, we briefly review some Radiomics approaches in breast cancer, focusing on the non-invasive distinction between in-situ and infiltrating breast tumors, and present a preliminary test using Radiomics signatures in DCE-MRI and machine learning, aimed to investigate the feasibility of distinguishing infiltrating cancer from ductal carcinoma in situ (DCIS) diagnosed by preoperative core needle biopsy.

Abbreviations: MRI: Magnetic Resonance Imaging; DCIS: Ductal Carcinoma In Situ; IDC: Invasive Ductal Carcinoma; DCE: Dynamic Contrast-Enhanced; ADC: Apparent Diffusion Coefficient; SVM: Support Vector Machine Algorithm; GLCM: Gray-Level Co-Occurrence Matrices; GLDM: Gray Level Dependence Matrices; GLSZM: Gray Level Size Zone Matrices; LOPO: Leave-One-Patient-Out

Introduction

Breast cancer is one of the leading causes of cancer-associated death among the female population worldwide [1]. In Italy, breast cancer affected about 52,000 new cases out of a total of 178,000 cases of all female cancers in 2018 [2]. Magnetic Resonance Imaging (MRI) is becoming more and more important in the clinical workflow of patients affected by breast carcinoma, because it enables the visual differentiation of normal tissues from pathological lesions owing to the increment of vascularity and capillary permeability of the latter [3-6]. Breast tumor can be classified into two broad types: in situ and invasive. The former is further subdivided into ductal and lobular, based on growth patterns and cytological characteristics. Ductal carcinoma in situ (DCIS) is more common than lobular carcinoma in situ (LCIS), accounting for 30-50\% of all mammography-detected breast cancers [7,8], and consists in neoplastic cells within the ductal epithelium of the breast. It normally does not infiltrate through the basal membrane. The most common malignant lesion...
is invasive ductal carcinoma (IDC) and accounts for approximately 70% of all malignant cases [9,10]. In recent years, there has been a decrease in number of deaths associated with breast cancer, due to earlier diagnosis, as well as to the introduction of advanced surgical techniques [11].

The identification of diagnostic and prognostic markers that enable the implementation of more targeted drug therapies remains a priority in the era of precision medicine [12]. Over the last years, the scientific community has been showing an increasing interest for the potentiality of quantitative imaging for clinical purposes, encouraged by the significant advancements within the medical image analysis field. This exponential interest led to the development of Radiomics, a new field of research that aims at the conversion of all the information contained in digital medical images into quantifiable features, and the subsequent mining of this data. These computational features, normally related to tumor size, shape, intensity, and texture, may be associated with clinical outcomes, genetic alterations and other characteristics of the lesion, defining what is called tumor Radiomics signature [13]. In this way, Radiomics seems able to offer imaging biomarkers useful to diagnosis and to predict the response to therapy and the risk of recurrence [14]. In this paper, a small review of the applications of Radiomics to breast cancer is given, particularly targeted at the non-invasive distinction between in-situ and infiltrative tumors, and the preliminary results of a limited case study are reported.

**Radiomics for Infiltrative vs In-Situ Distinction**

Only recently (mainly after 2015), Radiomics approaches were applied to breast cancer [15], with the majority of studies being published in 2017 [12]. Among these studies, Radiomics was mainly investigated with MRI and focused on the ability of predicting malignancy, response to neoadjuvant chemotherapy, prognostic factors, molecular subtypes and risk of recurrence [12,14,16]. Some recent studies addressed the distinction between in-situ and invasive breast cancer. For DCIS, upstaging to IDC at surgical excision occurs in roughly 25% of cases [17]. Failure to diagnose invasive cancer prior to surgery may have numerous implications. Normally, DCIS does not have metastatic potential. Thus, evaluation of regional or distant lymph nodes is usually not performed. Secondly, treatments are different between these two groups, so patients with IDC may need to undergo additional surgical procedures. This leads to the need to find different approaches to avoid unnecessary treatments in patients with non-invasive tumors, and many efforts should be made to achieve a diagnostic test for differentiation of in-situ from invasive breast cancer. Although a few studies examine a pharmacological intervention as solution [18,19], others would prefer the Watch & Wait approach instead of immediate surgery, which obviously avoids aggressive intervention [20].

In literature, there are still few reports that address the issue of characterizing invasive and non-invasive breast lesions. Drukker et al. [21] adopted a random forest classifier in a leave-one-out training/testing paradigm on Radiomics features extracted from dynamic contrast-enhanced (DCE-MRI) images (58 DCIS and 190 IDC) in the task of distinguishing between in situ and invasive breast cancer. They assessed the performance of the classifier by using the area under the receiver operating characteristic curve (AUC) which was 0.90. Li et al. [22] investigated the feasibility of predicting invasive breast cancer from in situ through a Radiomics approach on mammography: they extracted 569 Radiomics features from microcalcifications of 161 pure DCIS and 89 IDC, and evaluated various combinations of feature selection and classification methods. The optimal machine learning method was achieved using both Radiomics and routinely clinical imaging characteristics (AUC = 0.72). Another research group [23] addressed this problem by evaluating whether the apparent diffusion coefficient (ADC) extracted through diffusion-weighted MRI (DWI) could be used as a biomarker able to differentiate in situ from non-invasive DCIS. DWI measures the random movement of water molecules (i.e. Brownian movement) therefore depicting the diffusivity of the examined tissue, providing a surrogate marker for tissue micro-structure and densities of the cells [24,25]. The principle underlying the use of ADC to discriminate between in situ and invasive cancer, is that the latter spreads throughout the breast tissue by degrading tissue structure by means of proteolytic activity. Thus, tissue changes and chronic inflammatory reaction to proteolysis lead to a relative or absolute reduction of extracellular water content. What is then expected, is a reduction of ADC of invasive compared with non-invasive cancer. In order to prove the hypothesis, Bichel et al. [23] analyzed 21 DCIS and 155 IDC, finding ADC mean values significantly different between the two groups (p<0.001 and AUC = 0.89) [23]. Bhooshan et al. [26] analyzed DCE-MRI from 132 benign, 71 DCIS and 150 IDC in which they employed a Radiomics approach in order to discriminate in situ vs invasive breast tumors, but also between metastatic and non metastatic lesions, obtaining AUC = 0.83. In particular, they used combined computer-extracted MR imaging kinetic and morphologic features with the task of classifying between DCIS and IDC, and - within the invasive tumors - further classified into negative or positive axillary lymph node involvement. Finally, Zhe et al. [27] tried a deep learning approach on breast MRI for predicting of invasive disease following the diagnosis of DCIS. They adopted a transfer learning strategy, in which a pre-trained network (GoogleNet) was used on 131 DCIS images as a starting point followed by a deep feature based method, where the feature map of a certain layer of the pre-trained model was used as features to train a support vector machine algorithm (SVM), through a classical machine learning approach. They obtained AUC = 0.70, highlighting the fact that convolutional neural networks could potentially be used to predict DCIS upstaging.

**A Case Study**

In this section we report the preliminary results of a Radiomics investigation focused on the distinction between DCIS and IDC. The purpose was to determine the capability of machine learning to build statistics models for diagnosis, classification, and prediction based on Radiomics signatures in preoperative DCE-MRI.
We used a dataset of 30 anonymized DCE-MRI scans of breast cancer (25 DCIS and 5 IDC) acquired at the ‘Di Summa-Perrino’ Hospital of Brindisi (Italy). The MRI sequence used was dynamic eTHRIVE with fat suppression, on a Philips Achieva 1.5 T MRI equipment. A Region-Of-Interest (ROI) of the lesion was manually delimited slice per slice for each case by an expert radiologist in the post-contrast images. After 3D tumour segmentation, the images were resampled to isometric 1-mm pixel size and Radiomics features were extracted from the ROI. We computed 18 first order histogram features, and several textural features quantifying intra-tumor heterogeneity: 22 features from gray-level co-occurrence matrices (GLCM), 16 from gray-level run length matrices (GLRLM), 14 from gray level dependence matrices (GLDM), and 16 from gray level size zone matrices (GLSZM) [28]. Before classification we used recursive feature elimination to discriminate the redundant and irrelevant data. The Extreme Gradient Boosting (XGBoost) classifier (an implementation of gradient boosted decision trees) [29] was used in a leave-one-patient-out (LOPO) cross-validation scheme. At each iteration the features were normalized to [0,1] using min–max normalization on the training subjects and subsequently applying the calculated normalization parameters to each test patient features. Performance for the classification task was assessed using different metrics, such as average precision-recall, balanced accuracy, confusion matrix, Matthews correlation coefficient and AUC from ROC curve (Receiver Operating Characteristic). All hyperparameters of XGBoost classifier were optimized for our unbalanced dataset. Preprocessing, feature extraction and classification were implemented using python 3.7 and pyradiomics (https://pyradiomics.readthedocs.io/en/latest/index.html). The evaluation of the trained classifier reported an average precision-recall score of 0.38, a balanced accuracy score of 0.76, a Matthews correlation coefficient of 0.52 and a ROC curve with an AUC of 0.72. The model correctly classified 26 subjects.

Conclusion

The non-invasive, reliable, pre-operative distinction between infiltrative and in-situ breast cancer represents an important challenge in the biomedical field. The contribution reported in our preliminary monocentric work aims to provide an automated clinical diagnosis tool and shows a final balanced accuracy score of 0.76. Its main limitation consists in the small sample size and the obvious imbalance of diagnoses towards infiltrating breast tumors. In order to make the system able to generalize, and therefore to increase its quality, it is necessary to increase the size of the dataset, experimenting methods to make the dataset less unbalanced. In perspective, this result is expected to be achieved by involving different hospitals, thus creating a multicenter study.

Acknowledgement

This work was supported in part by the Trans Adriatic Pipeline project within the TAP Start funding programme.

Conflict of Interest

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

References


