Multicontrast MRI-based radiomics for the prediction of pathological complete response to neoadjuvant chemotherapy in patients with early triple negative breast cancer

Angeline Nemeth, Pierre Chaudet, Benjamin Leporq, Pierre-Etienne Heudel, Fanny Barabas, Olivier Tredan, Isabelle Treilleux, Agnès Coulon, Frank Pilleul, Olivier Beuf

Title page

Multicontrast MRI-based radiomics for the prediction of pathological complete response to neoadjuvant chemotherapy in patients with early triple negative breast cancer

Authors:

Angeline NEMETH, PhD,1 Pierre CHAUDET, MD,2 Benjamin LEPORQ, PhD,1 Pierre-Etienne HEUDEL, MD,3 Fanny BARABAS,2 Olivier TREDAN, MD-PhD,3 Isabelle TREILLEUX, MD-PhD,4 Agnès COULON, MD,2 Frank PILLEUL, MD-PhD,1,2 and Olivier BEUF, PhD,1

Author Affiliations:

1Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Etienne, CNRS, Inserm, CREATIS UMR 5220, U1206, F69621, Lyon, France

2Department of Radiology, Centre Léon Bérard, Lyon, France

3Department of Medical Oncology, Centre Léon Bérard, Lyon, France

4Department of Pathology, Centre Leon Bérard, Lyon, France

Corresponding author:

Benjamin Leporq, +33 4 69 85 62 58

Benjamin.leporq@creatis.insa-lyon.fr

Centre Léon Bérard, 28 Prom. Léa et Napoléon Bullukian, 69008 Lyon, France

Acknowledgments

This study was conducted as part of the LABEX PRIMES (ANR-11-LABX-0063) of the "Université de Lyon", within the "Investissements d'Avenir" program (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR). This study was also supported by the SIRIC LyriCAN grant (INCa_INSERM_DGOS_12563). We thank Sophie Darnis for her help with English language editing.

Abstract

Introduction:

To assess pre-therapeutic MRI-based radiomic analysis to predict the pathological complete response to neoadjuvant chemotherapy (NAC) in women with early triple negative breast cancer (TN).

Materials and methods:

This monocentric retrospective study included 75 TN female patients with MRI (T1-weighted, T2-weighted, diffusion-weighted and dynamic contrast enhancement images) performed before NAC. For each patient, the tumor(s) and the parenchyma were independently segmented and analyzed with radiomic analysis to extract shape, size, and texture features. Several sets of features were realized based on the 4 different sequence images. Performances of 4 classifiers (random forest, multilayer perceptron, support vector machine (SVM) with linear or quadratic kernel) were compared based on pathological complete response (defined on the excised tissues), on 100 draws with 75% as training set and 25% as test.

Results:

The combination of features extracted from different MR images improved the classifier performance (more precisely, the features from T1W, T2W and DWI). The SVM with quadratic kernel showed the best performance with a mean AUC of 0.83, a sensitivity of 0.85 and a specificity of 0.75 in the test set.

Conclusion:

MRI-based radiomics may be relevant to predict NAC response in TN cancer. Our results promote the use of multi-contrast MRI sources for radiomics, providing enrich source of information to enhance model generalization.

Keywords

Breast Cancer; Radiomics; Multi-contrast MRI; Triple Negative Breast Cancer.

Abbreviations

- AUC: Area Under the Curve ROC
- DCE: dynamic contrast enhancement

- HER: Human Epidermal Growth Factor
- VOI: volume of interest
- MLP: multilayer perceptron
- MRI: magnetic resonance imaging
- NAC: neo-adjuvant chemotherapy
- pCR: pathological complete response
- ROC: receiver operating characteristic
- SVM: support vector machine
- TN: triple negative
- TNM: tumor, node, metastases
- T1W: T1-weighted imaging
- T2W: T2-weighted imaging
- DWI: diffusion weighted imaging
- GLCM: Gray-level co-occurrence matrix
- GLSZM: Gray-level size zone matrix
- NGTDM: Neighborhood gray tone difference matrix
- SURF: Speed-Up Robust Features
- SUB3: subtraction between the image 3 min post-injection and the image pre-injection

TEXT

INTRODUCTION

Breast cancer is the most frequently diagnosed cancer in women with 2,090,000 new cases and 627,000 deaths worldwide in 2018 [1]. Among the different types of breast cancer, triple negative (TN) cancer is characterized by estrogen and progesterone receptor level lower than 10%, and an absence of over-expression of HER-2 (Human Epidermal Growth Factor Receptor-2). TN cancers account for 10 to 24% of all breast cancers, and 57 to 88% of cancers with *BRCA1* mutation in women [2]. TN tumors are generally larger, diagnosed at the highest grade, and associated with worse prognosis [3]. At early stage, patients with TN cancer receive systemic treatments generally limited to cytotoxic chemotherapy, no targeted therapies are currently proposed. Neoadjuvant chemotherapy (NAC) treatment are used: i) to reduce the initial tumor volume in order to allow a conservative surgical treatment; ii) to better eradicate the micrometastatic disease; iii) to assess tumor chemo-sensibility to determine the most appropriate adjuvant treatments [4]. The pathological complete response (pCR) to NAC is an important prognostic factor for the disease-free survival and overall survival in breast cancer [5, 6]. The efficacy of NAC varies according to the tumor's genetic profile and pCR levels from 10 to 50% [7]. Therefore, there is a clinical need to identify patients who will not respond to NAC to eventually direct them to alternative therapeutic strategies.

According to the latest EUSOBI recommendations [8], pre- and post-chemotherapy MRI examination should be performed in women with early breast cancer receiving NAC. Since 2012 [9], a new discipline called "radiomics" has drawn increasing attention in cancer research in disease detection, diagnosis, and prediction of treatment response, and several studies investigated prediction of pCR to breast cancer chemotherapy [10–14]. Radiomics is based on the hypothesis that genetics, molecular, cellular and tissular modifications can be observed on images [15,16]. Technically, radiomics consist in extracting a high number of quantitative parameters from radiologic images in order to determine their relationships with the underlying pathophysiology [17,18]. Moreover, through functional and anatomical information that MRI provided with on whole tissues, MRI-based radiomics allows to access to quantitative information refining the entire tumor and its micro-environment, and to probe tumor heterogeneity [19]. Some radiomic studies focused not only on the tumor, but also on the surrounding mammary parenchyma. The tumor microenvironment is known to partly contribute to the

progression of breast cancer [20]. The radiomic analysis of the pre-tumoral environment was shown to be as important as the tumor analysis itself [14].

In this study, we explored the performance of radiomics in tumor and ipsilateral parenchymal mammary MRI to predict the pCR to NAC in TN breast cancer female patients.

METHOD

Study design

Patients enrolled in this retrospective study underwent a pre-therapeutic MRI protocol between January 2008 and Jun 2017 in the French comprehensive cancer center "Centre Léon Bérard". All patients had an early triple negative breast cancer (i.e. without metastasis) and were treated with neoadjuvant chemotherapy (NAC) before a surgical treatment. All women received a sequential NAC therapy based on anthracyclines-cyclophosphamide, then taxanes. Data were excluded for: i) poor quality of MRI imaging (moving artifacts, missing of some images); ii) NAC treatment initiated prior to MRI. This retrospective study was approved by our institutional review board and the requirement to obtained informed consent was waived.

Pre-therapeutic MRI protocol

All breast MRI examinations were performed at Leon Berard center with patient in prone position using a 1.5T Achieva system (Philips Healthcare, Best, Netherlands) and with the use of a dedicated seven-channel or sixteen-channel breast surface coil (SENSE-Breast-7 or SENSE-Breast16M). The pre-therapeutic MRI protocol was sequentially composed of a T1-weighted imaging, a T2-weigthed imaging, a diffusion weighted imaging and a dynamic contrast enhanced imaging (Figure 1). MRI protocol parameters are summarized in Table 1. Dynamic contrast enhancement (DCE) MR imaging was acquired after intravenous injection of 0.1 mmol/kg gadolinium-based contrast agent. Images noted "SUB3" were obtained subtracting images acquired 3 min after the injection and those acquired before injection using DCE-MRI.

Segmentation of volume of interest

The 3D volume of interest (VOI) was delineated manually using itk-SNAP software

(www.itksnap.org) on few slices of SUB3 images; the inter-slice interpolation option was used to complete the mask between slices to have a 3D volume. More precisely, for each patient, tumor and parenchyma were initially segmented by a radiologist intern PC, after correcting by a senior radiologist (more than 10 years of experience in breast imaging AC). In the case of multiple lesions, each of them was considered independently with separate 3D binary mask. The delineation of the parenchyma included the fibroglandular tissue and the adipose tissue and excluded the skin and the tumor. The segmentations of tumor and parenchyma resulting from the consensus of the two radiologists PC and AC were then used for the extraction of size, shape, intensity distribution, and other texture features. Areas impacted by the presence of a clip were excluded from the VOI for the computation of intensity distribution and texture features (but not for shape and size features). An affine transformation was used to reposition VOIs segmented on SUB3 images on the other images (T1W, T2W and DWI) using matrix dimension and patient position information. A third expert FB was asked to delineate a subset of twenty-four tumors in order to analyze the variability of segmentations with the difficulty to have any access to the assessment of the other two radiologists and without access to the patient file.

Features extraction

We tested multiple configurations. We first use only SUB3 images with features extracted from lesion and parenchyma, and progressively include features from T1-weitghted images, T2-weitghted images, and DW images in the feature set. Extraction of feature set was performed with MATLAB 2019a (The Mathworks, Natick, MA, USA) using an in-house software. 3D-image intensities were filtered using Collewet method [21]. The initial feature set of a ROI was composed of 342 features: 96 shape and size characteristics, 14 intensity distribution characteristics, and 232 textural characteristics. As shown in the Figure 2, size and shape features were directly extracted from the binary masks and were based on morphological skeletonization and distance transforms, affine moment invariants [22], Hu moment [23], Zernike moment [24, 25], and conventional metrics. Intensity distribution features were extracted from masked MR images from the histogram built with 256 bins (14 features: average, standard deviation, full width at half maximum, variance, minimum, maximum, range, interquartile range, kurtosis, skewness, entropy energy, root mean square, mean deviation, median deviation). Before the

extraction of texture features, voxels were isotropically resampled using an affine transformation and a nearest-neighbor interpolation and then discretized in a smaller number of gray levels. This operation was done using an equal probability algorithm to define decision thresholds in the volume such as the number of voxels for a given reconstructed level is the same in the quantized volume for all gray levels. Images were discretized in 8, 16, 24, 32, 48 and 64 grey levels and for each level four matrix were built: GLCM (Gray-level co-occurrence matrix) (n=21), GLRLM (Gray-level run length matrix) (n=13), GLSZM (Gray-level size zone matrix) (n=13) and NGTDM (Neighborhood gray tone difference matrix) (n=5) from which characteristics were extracted. Frequency domain-based texture features were extracted from the Gabor filters responses and from features extracted from image spectrum after 2D discrete Fourier transform. GLCM and GLRLM will be computed for 4 directions (0°, 45°, 90° and 135°) with an offset of 1 pixel. For GLSZM and NGTDM, a 26-pixel connectivity will be used. For Gabor filtering, 5 scales, 6 orientations, and a minimal wavelength of 3 were used. Other texture feature based on images primitive were also extracted using different detectors and descriptors (such as Speed-Up Robust Features (SURF) detector, Local Oriented Statistics Information Booster (LOSIB) descriptor, Harris detector); lacunarity computation or quad tree decomposition.
These features were already described in previous studies [10,11,26]. For multiparametric case, only textural and intensity features (set of 492 features) were computed on the other imaging (T1W, T2W and DWI) and added to the initial feature set as illustrated in Figure 2.

Data mining

The test set included 25% of the total number of tumors, randomly selected from the whole data set, with a balance between pCR and non-pCR. One hundred different configurations of the training and test sets were used. Z-score normalization was applied on each features of the feature set. Then, a dimension reduction was applied using ReliefF method [27] to select the twenty most relevant features on the whole data set. From the reduced feature set, supervised machine learning was used to build the prediction model. Four classifiers were evaluated: a multilayer perceptron (MLP) trained with a stochastic gradient algorithm using an adaptive learning rate and a regularization of the synaptic weigths (= 0.1, 5 mini-batches, 30 hidden nodes, and 60 epochs); a support vector machine (SVM) with a linear kernel; a SVM with a quadratic kernel (trained with box-constraint c = 1) and a random

forest (3 splits and 50 learning cycles). The difference between the AUC for the training set and for the test set was used to evaluate the overfitting of the classification method. We evaluated the performance of the classification models thanks to the area under the curve (AUC) of the receiver operating characteristic (ROC). The difference in AUC between the training and the testing sets was used as an indicator for the predictive model to generalize the estimation.

Statistical Analysis

A Sørensen–Dice index was computed to compared the delineation given by various radiologists. An intraclass correlation coefficient (ICC) for each radiomic features was computed using package psych in R software. As we compared the radiomic features from three segmentations (initial P.C, corrected by A.C. and independent F.B.), we used the "ICC2" definition (this measure is one of absolute agreement in the ratings and could be generalize to other observers).

A linear mixed regression model with random intercept was used to evaluate the effect of the multi-contrast feature sets, the effect of the choice of classifier, and the effect of additive information from the parenchyma in the feature set on the AUC values (details in supplementary information). T-test was performed for specific pairwise comparison. Analyses were made using R software, v. 3.6.1 (Vienna, Austria).

Pathological examination and response to treatment

The breast cancer was diagnosed on biopsy. All anatomopathological diagnostics were realized by expert pathologists specialized in breast cancer from our institution. The expression of estrogen receptor, progesterone receptors, and HER-2 status was determined on histopathological pre-therapeutic biopsy samples. Hormone receptor negative status was defined if less than 10% of cells revealed staining for estrogen and progesterone receptors. The expression of HER-2 was considered negative if lower than 1+ in immunohistochemistry. Tumors with a score at 2+ required additional in situ hybridization to determine the amplification or non-amplification of HER-2. After a 6-month NAC treatment, all patients underwent breast surgical intervention (lumpectomy or mastectomy). The complete pathological response was defined by the absence of invasive residual tumor in the resected tissues (carcinoma in situ could be found) and the absence of axillary nodal metastasis. The TNM

stage was ypT0 ypNO or ypTis ypNO, according to 2012 seventh editions of the AJCC Cancer Staging Manual 2012[28].

RESULTS

Clinical characteristics

Among 79 patients eligible for study enrolment, four patients were excluded: one for MRI-protocol performed one week after the treatment initiation, and three because the SUB3 images were missing or not complete after the data transfer from the archive server. Among the 75 patients enrolled, 11 patients presented two lesions, and 3 patients presented three lesions. Two patients presented bilateral tumors. Lesions were analyzed independently. 14 out of these 92 tumors had a clip. Three lesions were excluded for inadequate size to perform radiomic processing. Tumor characteristics are summarized in Table 2.

8 out of the 75 patients underwent MRI protocol using the SENSE-Breast-7 coil, and the remaining patients the SENSE-Breast16M. No significant difference in signal-to-noise ratio was observed between images acquisitions with the SENSE-Breast-7 and the SENSE-Breast16M coils (t-test, p-value = 0.33). For some patient, images from the pre-therapeutic MRI images were missing; details are summarized in Table 3 and reported the proportion of pCR and non-pCR.

Comparison of multi-contrast imaging features

Out of the 89 tumors analyzed, 70 used the whole MRI protocol (T1W, T2W, DWI, and DCE). Multiple combinations of feature set were tested to evaluate the necessity of multiple contrasts imaging (Figure 3). Moreover, different classifiers were used and their performances were compared.

Based on the AUC in the training set, the SVM with a quadratic kernel showed the best results (mean AUC of 0.99, Figure 3J). Random Forest and Multilayer perceptron also provided very good results with a mean AUC of 0.93 and 0.90, respectively, for the training set. The “SVM order 1” classifier showed the smallest difference in AUC between the training and test sets reflecting the smallest overfitting (Figure 3 C, F, I, L), or largest bias.

The random forest, “SVM order 2”, and the MLP classifiers suggest advantages in adding information from different images issued from different MRI sequences. The most interesting configuration of

feature set appeared to be the combination of DWI, T1W, and T2W features: firstly, the AUC in the validation set was larger than in other configurations even considering the configuration with only T1W features (+ 0.034, p <0.001, linear mixed model with Dunnett post-hoc test, details in supplementary information); secondly, the overfitting expressed by the difference in AUCs was in the same range than that of the "SVM order 1" (+0.008, p =0.564, t-test).

The same parameters were used without any characteristics of parenchyma while keeping selected the 20 most relevant parameters. A significant decreased in AUC for the test set was observed for "SVM order 2" (- 0.047, p <0.001, linear mixed model with Dunnett post-hoc test, details in supplementary information). These results showed that the characteristics of both parenchyma and tumors seemed to allow a better prediction of pCR. Unfortunately, the AUC results depended on the repartition of the data in the training set and in test set, and the ranges of AUCs in the test set were large.

To go further, it would be interesting to accurately identify the non-pCR patients to avoid exposure to ineffective systemic treatment. In that case, the specificity of the predictive model was the most important. We observed that sensitivity was higher than specificity in our series (Figure 4).

Effect of features selected

The "SVM order 2" combined with DWI, T1, and T2 features seemed to better predict pCR. Therefore, we analyzed the effect of the number of features selected before applying classification. The ReliefF method allowed to promptly classify the most independent features (Table 4). The textural features appear more relevant than shape and 1er order textural features (26 textural features against 4 shape feature). The Figure 5 shows: i) an increase in AUC, specificity and sensitivity values from 1 to 9 parameters; ii) a plateau from 9 to 24 parameters; iii) a slight decrease after 24 parameters. The nine most relevant features appeared as sufficient. In relevant shape feature, the one from parenchyma was very interesting because it reflects the proportion of the sane parenchyma against the tumor (the tumor area was removed from the parenchyma delineation in the binary mask). For the T1W without fat suppression, lobules, ducts and bloods vessels appear in lower values than fatty tissue. The features Short-Run Low Gray-level is the distribution of the short homogeneous runs with low grey-levels and appeared to be a relevant textural feature from parenchyma on T1W imaging.

Variability of intra-observer delineation and radiomic feature implications

The variability analyses were done on a subset of 24 lesions. The segmentation approved by the radiologist AC was considered the reference. A mean DICE coefficient of 0.95 ± 0.07 (range of 0.74 - 1.00) reflecting the concordance between the two radiologists (PC and AC). The mean DICE coefficient was 0.84 ± 0.10 (range of 0.56 - 0.99) comparing the delineation approved by AC and FB. The worst result (DICE = 0.56) came from a stack of images where the contrast enhancement was very low and therefore difficult to segment without information on the pathology. The majority of DICE coefficient was over 0.80. To measure the impact of delineation variation on radiomic features, an ICC was computed for each feature. On the most relevant features highlighted by the ReliefF method in our study, we observed that textural features have a good ICC (mean 0.88 range 0.65-0.98) but the shape features have a worst ICC2 (mean 0.44 range 0.31-0.56). The textural features represented 75% of the nine most relevant features.

DISCUSSION

This study shows that multi-contrast MRI-based radiomics allows to predict the neo-adjuvant chemotherapy response in patients with early triple negative breast cancer with good performances.

This study also showed that the choice of classifier strongly influences model performances. While SVM with a linear kernel generalized well, probably due to lower variance and higher bias (highlighted by lower AUC in the training set), other non-linear classifiers such as neural network, quadratic-kernel SVM, and Random Forest, increase performances by reducing bias effect (high performances in the training set), and increase variance and therefore overfitting, as indicated by increased AUC differences between dataset. However, the increase in variety of processable data combining different sources of radiomics (multi-contrast MRI) reduced variance effect as evidenced by reduced AUC differences between dataset when radiomics aggregated from different contrasts. Our results showed that the best model (AUROC at 0.83) was obtained with SVM with a quadratic kernel trained from aggregated radiomics extracted from T1, T2, and diffusion weighted images.

From a clinical point of view, radiomic signature could help to better predict pathological complete

response, and enhanced stratification of patients with an excellent prognosis, and patients at high risk of chemoresistance. In the latter population, systemic treatment combined to the standard NAC (anthracycline/taxane) may be added, such as platinum-based NAC [29], a PARP inhibitor [30], or even an immune-checkpoint-inhibitor. Indeed, the KEYNOTE-522 trial [31] investigating the addition of prembrolizumab to neo-adjuvant standard chemotherapy showed an increase of 64.8% of pathological complete response, compared with 51.2% in the placebo control group. Therefore, it seems important to appropriately select patients who may benefit the most from appropriate systemic treatment, highlighting the medical need for these non-invasive stratification methods.

The most informative sequences identified were the association of T1/T2 morphological sequences and diffusion sequence. However, these results also depend on the DCE-MRI sequence used for the segmentation of the lesion. Indeed, we observed some pixels in the fat surrounding the tumors in the mask registered on T1W images; which would not have been integrated if the segmentation had been directly performed on T1W images. The parenchyma features appeared to be relevant in this study. The shape feature computed on the parenchyma binary mask (the difference between the parenchyma and the tumor delineation) appeared to be more relevant than the shape feature computed on tumor binary mask. More generally, the most relevant features were high order of textural features. These textural features would reflect the textural complexity from a macroscopic point of view but the tumor heterogeneity in a microscopy point view. Moreover, the textural features appeared to be more robust to inter-observer variability (different delineation) with high intraclass correlation coefficient.

Our results are consistent with previous studies showing that AUC ranged from 0.67 to 0.87 [10–12,14], allow to predict the response to breast cancer chemotherapy. Among these studies, Braman and collaborators showed that the triple negative subgroups had the best results with a 0.93 AUC, and Liu and colleagues reported AUC at 0.86 with fewer TN patients. Our series showed that 64% of the patients achieved a pathological complete response; this rate is higher than that reported in previous studies with 31% of patients complete pathological responses in [32].

Our study has some limitations. First, the size of our population was limited to 75 patients and the multiparametric radiomic analyses were only achieved in 70 lesions. However, to the best of our knowledge, this TN cohort was the largest cohort dedicated to radiomic studies. The lesion

segmentation was done manually, which may introduce a degree of subjectivity. In addition, this study was performed in a single institution study; and extrapolating results to population of another center is limited. Lastly, MRI protocols are not standardized between centers in the field of breast cancer analyses inducing variability in the radiomic feature sets. Several studies have shown the sensitivity of radiomic parameters to machine's change or reconstruction parameters [33–35].

Those results have to be confirmed in an external validation cohort with multicentric and prospective data, and more patients.

To conclude, our results confirm that MRI-based radiomics may be relevant to predict neoadjuvant chemotherapy response in early triple negative breast cancer. In addition, our results highlight the interest of using multi-contrast MRI as sources of radiomics to improve model generalization thanks to increased information variety.

Authors' Contribution

Nemeth: Study conception and design, Analysis and interpretation of data, Drafting of manuscript, Critical revision

Chaudet: Study conception and design, Acquisition of data, Analysis and interpretation of data, Drafting of manuscript

Leporq: Study conception and design, Analysis and interpretation of data, Critical revision

Heudel: Study conception and design, Acquisition of data

Barabas: Acquisition of data, Analysis and interpretation of data

Tredan: Study conception and design

Treilleux: Study conception and design

Coulon: Study conception and design, Acquisition of data, Analysis and interpretation of data, Drafting of manuscript

Pilleul: Analysis and interpretation of data, Critical revision

Beuf: Analysis and interpretation of data, Critical revision

Declarations

Ethics approval

Ethical approval was waived by the local Ethics Committee of our institution in view of the retrospective nature of the study and all the procedures being performed were part of the routine care.

REFERENCES

1. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, Znaor A, Bray F (2018) Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. Int J Cancer ijc.31937.

2. Billar JAY, Dueck AC, Stucky C-CH, Gray RJ, Wasif N, Northfelt DW, McCullough AE, Pockaj BA (2010) Triple-Negative Breast Cancers: Unique Clinical Presentations and Outcomes. Ann Surg Oncol 17:384–390.

3. Foulkes WD, Smith IE, Reis-Filho JS (2010) Triple-Negative Breast Cancer. New England Journal of Medicine 363:1938–1948.

4. Derks MGM, van de Velde CJH (2018) Neoadjuvant chemotherapy in breast cancer: more than just downsizing. The Lancet Oncology 19:2–3.

5. Kong X, Moran MS, Zhang N, Haffty B, Yang Q (2011) Meta-analysis confirms achieving pathological complete response after neoadjuvant chemotherapy predicts favourable prognosis for breast cancer patients. European Journal of Cancer 47:2084–2090.

6. Liedtke C, Mazouni C, Hess KR, André F, Tordai A, Mejia JA, Symmans WF, Gonzalez-Angulo AM, Hennessy B, Green M, Cristofanilli M, Hortobagyi GN, Pusztai L (2008) Response to Neoadjuvant Therapy and Long-Term Survival in Patients With Triple-Negative Breast Cancer. JCO 26:1275–1281.

7. Earl H, Provenzano E, Abraham J, Dunn J, Vallier A-L, Gounaris I, Hiller L (2015) Neoadjuvant trials in early breast cancer: pathological response at surgery and correlation to longer term outcomes – what does it all mean? BMC Med 13:234.

8. for the European Society of Breast Imaging (EUSOBI), with language review by Europa Donna–The European Breast Cancer Coalition, Mann RM, Balleyguier C, Baltzer PA, Bick U, Colin C, Cornford E, Evans A, Fallenberg E, Forrai G, Fuchsjäger MH, Gilbert FJ, Helbich TH, Heywang-

Köbrunner SH, Camps-Herrero J, Kuhl CK, Martincich L, Pediconi F, Panizza P, Pina LJ, Pijnappel RM, Pinker-Domenig K, Skaane P, Sardanelli F (2015) Breast MRI: EUSOBI recommendations for women's information. Eur Radiol 25:3669–3678.

9. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, Zegers CML, Gillies R, Boellard R, Dekker A, Aerts HJWL (2012) Radiomics: Extracting more information from medical images using advanced feature analysis. European Journal of Cancer 48:441–446.

10. Liu Z, Li Z, Qu J, Zhang R, Zhou X, Li L, Sun K, Tang Z, Jiang H, Li H, Xiong Q, Ding Y, Zhao X, Wang K, Liu Z, Tian J (2019) Radiomics of Multiparametric MRI for Pretreatment Prediction of Pathologic Complete Response to Neoadjuvant Chemotherapy in Breast Cancer: A Multicenter Study. Clin Cancer Res 25:3538–3547.

11. Chamming's F, Ueno Y, Ferré R, Kao E, Jannot A-S, Chong J, Omeroglu A, Mesurolle B, Reinhold C, Gallix B (2018) Features from Computerized Texture Analysis of Breast Cancers at Pretreatment MR Imaging Are Associated with Response to Neoadjuvant Chemotherapy. Radiology 286:412–420.

12. Fan M, Wu G, Cheng H, Zhang J, Shao G, Li L (2017) Radiomic analysis of DCE-MRI for prediction of response to neoadjuvant chemotherapy in breast cancer patients. European Journal of Radiology 94:140–147.

13. Weber JJ, Jochelson MS, Eaton A, Zabor EC, Barrio AV, Gemignani ML, Pilewskie M, Van Zee KJ, Morrow M, El-Tamer M (2017) MRI and Prediction of Pathologic Complete Response in the Breast and Axilla after Neoadjuvant Chemotherapy for Breast Cancer. Journal of the American College of Surgeons 225:740–746.

14. Braman NM, Etesami M, Prasanna P, Dubchuk C, Gilmore H, Tiwari P, Plecha D, Madabhushi A (2017) Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. Breast Cancer Res 19:57.

15. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, Sanduleanu S, Larue RTHM, Even AJG, Jochems A, van Wijk Y, Woodruff H, van Soest J, Lustberg T, Roelofs E, van Elmpt W, Dekker A, Mottaghy FM, Wildberger JE, Walsh S (2017) Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 14:749–762.

16. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: Images Are More than Pictures, They Are

Data. Radiology 278:563–577.

17. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, Sun K, Li L, Li B, Wang M, Tian J (2019) The Applications of Radiomics in Precision Diagnosis and Treatment of Oncology: Opportunities and Challenges. Theranostics 9:1303–1322.

18. Limkin EJ, Sun R, Dercle L, Zacharaki EI, Robert C, Reuzé S, Schernberg A, Paragios N, Deutsch E, Ferté C (2017) Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. Annals of Oncology 28:1191–1206.

19. Agner SC, Rosen MA, Englander S, Tomaszewski JE, Feldman MD, Zhang P, Mies C, Schnall MD, Madabhushi A (2014) Computerized Image Analysis for Identifying Triple-Negative Breast Cancers and Differentiating Them from Other Molecular Subtypes of Breast Cancer on Dynamic Contrast-enhanced MR Images: A Feasibility Study. Radiology 272:91–99.

20. Quail DF, Joyce JA (2013) Microenvironmental regulation of tumor progression and metastasis. Nat Med 19:1423–1437.

21. Collewet G, Strzelecki M, Mariette F (2004) Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. Magnetic Resonance Imaging 22:81–91.

22. Suk T, Flusser J (2003) Combined blur and affine moment invariants and their use in pattern recognition. Pattern Recognition 36:2895–2907.

23. Ming-Kuei Hu (1962) Visual pattern recognition by moment invariants. IEEE Trans Inform Theory 8:179–187.

24. Tahmasbi A, Saki F, Shokouhi SB (2011) Classification of benign and malignant masses based on Zernike moments. Computers in Biology and Medicine 41:726–735.

25. Saki F, Tahmasbi A, Soltanian-Zadeh H, Shokouhi SB (2013) Fast opposite weight learning rules with application in breast cancer diagnosis. Computers in Biology and Medicine 43:32–41.

26. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebers F, Rietbergen MM, Leemans CR, Dekker A, Quackenbush J, Gillies RJ, Lambin P (2014) Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 5:4006.

27. Kononenko I, Simec E, Robnik-Sikonja M Overcoming the myopia of inductive learning algorithms with RELIEFF. 17.

28. Compton CC, Byrd DR, Garcia-Aguilar J, Kurtzman SH, Olawaiye A, Washington MK (2012) AJCC Cancer Staging Atlas: A Companion to the Seventh Editions of the AJCC Cancer Staging Manual and Handbook, 2nd ed. doi: 10.1007/978-1-4614-2080-4

29. Poggio F, Bruzzone M, Ceppi M, Pondé NF, La Valle G, Del Mastro L, de Azambuja E, Lambertini M (2018) Platinum-based neoadjuvant chemotherapy in triple-negative breast cancer: a systematic review and meta-analysis. Annals of Oncology 29:1497–1508.

30. Loibl S, O'Shaughnessy J, Untch M, Sikov WM, Rugo HS, McKee MD, Huober J, Golshan M, von Minckwitz G, Maag D, Sullivan D, Wolmark N, McIntyre K, Ponce Lorenzo JJ, Metzger Filho O, Rastogi P, Symmans WF, Liu X, Geyer CE (2018) Addition of the PARP inhibitor veliparib plus carboplatin or carboplatin alone to standard neoadjuvant chemotherapy in triple-negative breast cancer (BrighTNess): a randomised, phase 3 trial. The Lancet Oncology 19:497–509.

31. Schmid P, Cortes J, Pusztai L, McArthur H, Kümmel S, Bergh J, Denkert C, Park YH, Hui R, Harbeck N, Takahashi M, Foukakis T, Fasching PA, Cardoso F, Untch M, Jia L, Karantza V, Zhao J, Aktan G, Dent R, O'Shaughnessy J, KEYNOTE-522 Investigators (2020) Pembrolizumab for Early Triple-Negative Breast Cancer. N Engl J Med 382:810–821.

32. Houssami N, Macaskill P, von Minckwitz G, Marinovich ML, Mamounas E (2012) Meta-analysis of the association of breast cancer subtype and pathologic complete response to neoadjuvant chemotherapy. European Journal of Cancer 48:3342–3354.

33. Yan J, Chu-Shern JL, Loi HY, Khor LK, Sinha AK, Quek ST, Tham IWK, Townsend D (2015) Impact of Image Reconstruction Settings on Texture Features in 18F-FDG PET. Journal of Nuclear Medicine 56:1667–1673.

34. Fortin J-P, Parker D, Tunç B, Watanabe T, Elliott MA, Ruparel K, Roalf DR, Satterthwaite TD, Gur RC, Gur RE, Schultz RT, Verma R, Shinohara RT (2017) Harmonization of multi-site diffusion tensor imaging data. NeuroImage 161:149–170.

35. Reuzé S, Orlhac F, Chargari C, Nioche C, Limkin E, Riet F, Escande A, Haie-Meder C, Dercle L, Gouy S, Buvat I, Deutsch E, Robert C (2017) Prediction of cervical cancer recurrence using textural features extracted from 18F-FDG PET images acquired with different scanners. Oncotarget. doi: 10.18632/oncotarget.17856

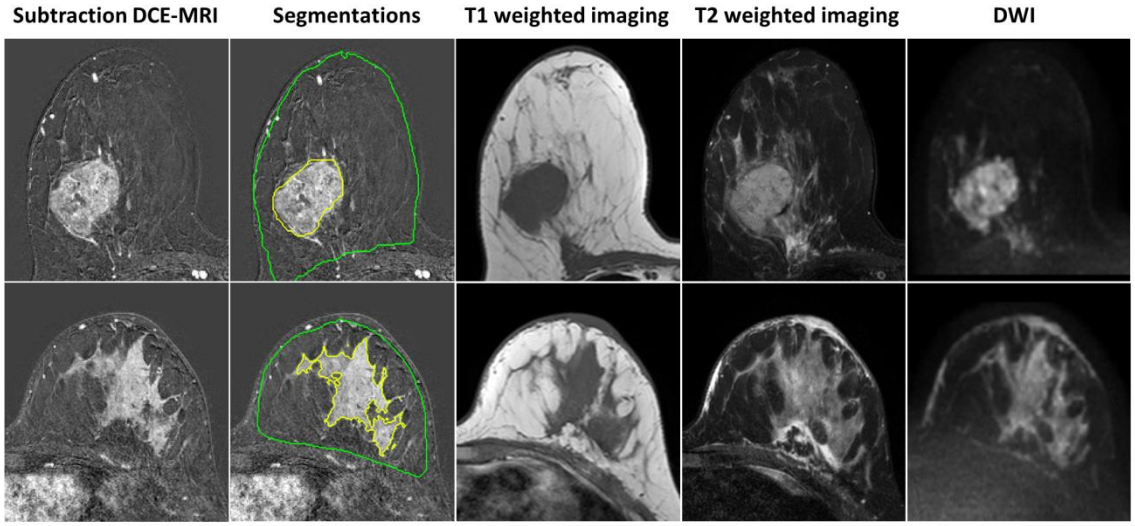


Figure 1. Two triple-negative breast cancers imaged with pre-therapeutic MRI protocol. First line: a mass tumor is shown; second line: a non-mass tumor is illustrated. Segmentation of tumor (in yellow) and segmentation of parenchyma (in green excluding tumor area) were made on derived DCE-MRI. For non-mass tumor, the option "snake" of ITK-SNAP was used to perform the segmentation.

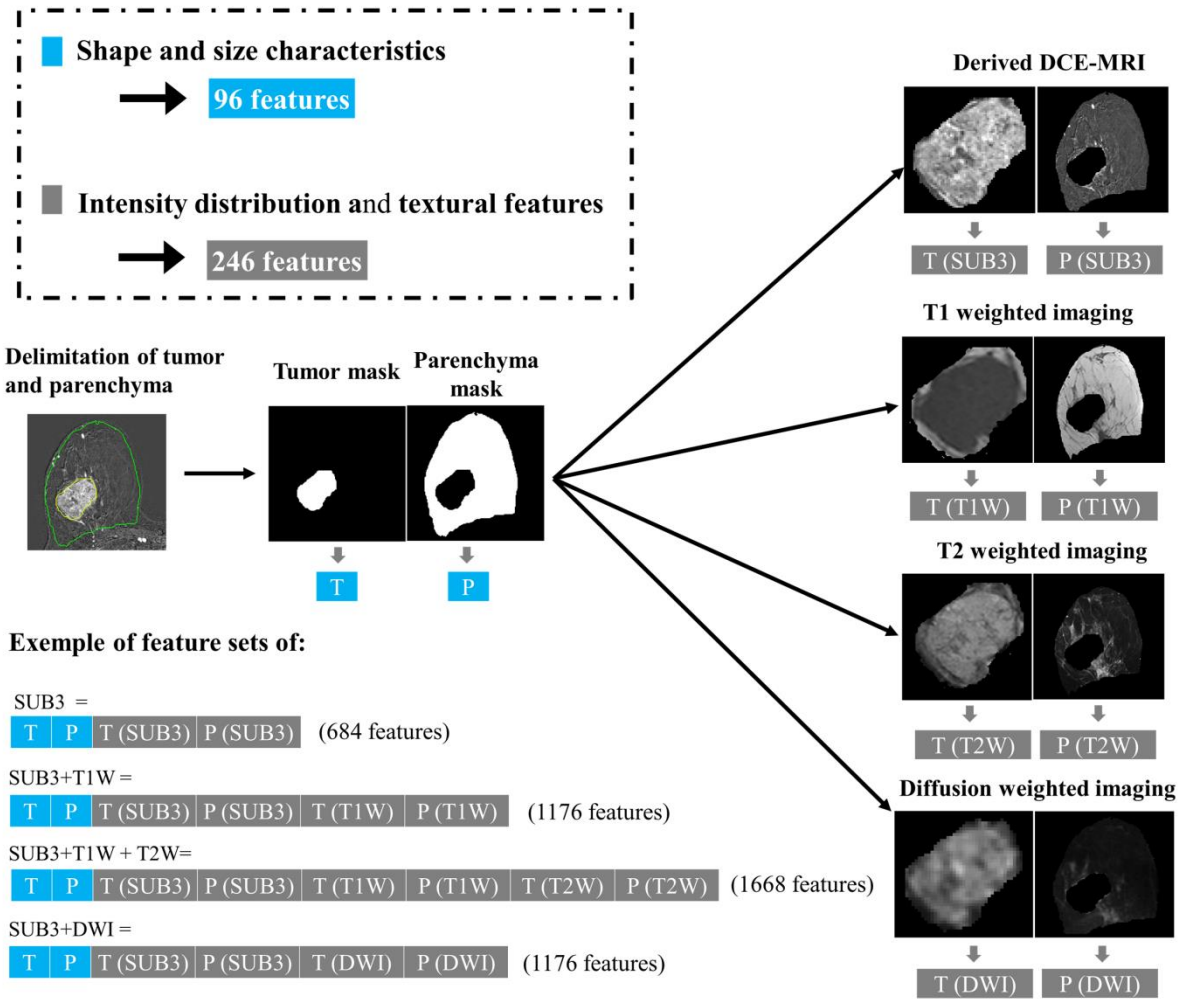
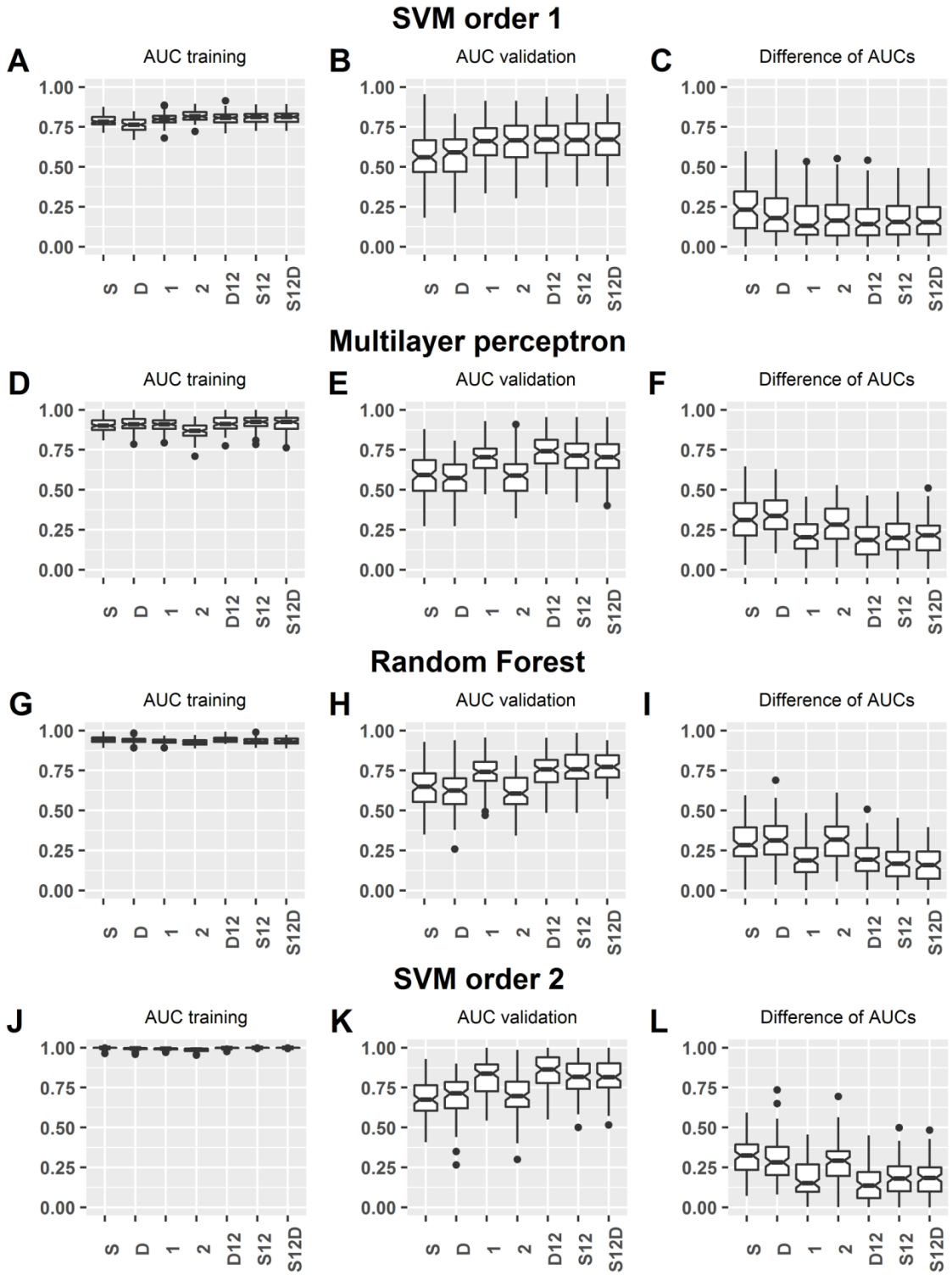


Figure 2. Pipeline of radiomic feature extraction and combination of feature sets.



488

489 **Figure 3. Boxplots of different configurations of radiomic feature set.** AUC: area under the curve
490 ROC for the training set and for the test set. For the case "S" only features from the SUB3 were
491 included in the feature set; "D" corresponds to the diffusion weighted imaging, "1" corresponds to the
492 T1-weighted imaging and "2" corresponds to the T2-weighted imaging. The results of all possible
493 configurations are presented as supplementary information.

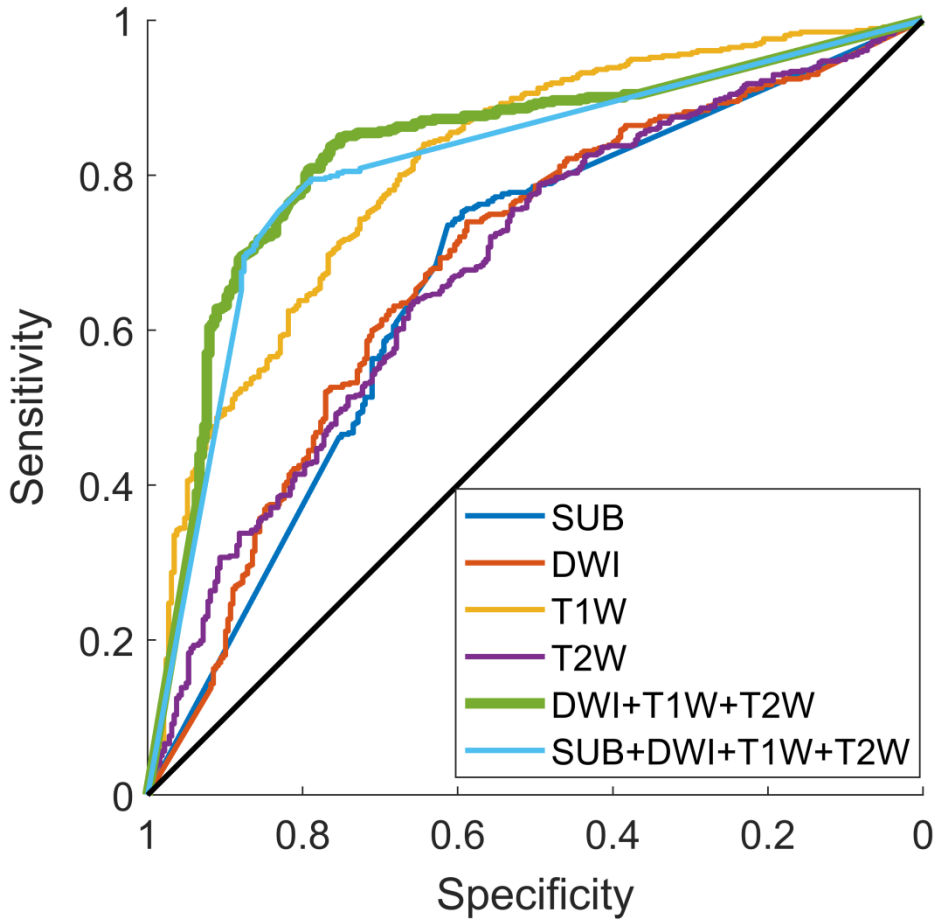
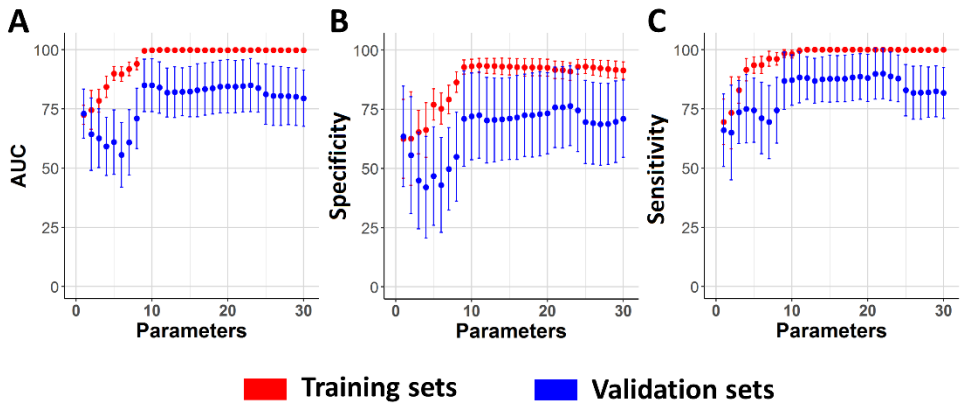


Figure 4: ROC curves for the SVM classifier with quadratic kernel. The most interesting configuration of feature set appeared to be the combination of DWI, T1W and T2W features with AUC =0.83, sensitivity =0.85 and specificity =0.75.

499



500

501 **Figure 5: Variation of model performances in training and test sets according to the number of**

502 **feature selected with the ReliefF method.**

503

Table 1: Summary of sequence parameters of T1W (T1-weigthed imaging), T2W (T2-weighted imaging), DWI (diffusion weighted imaging) and DCE (dynamic contrast enhancement).

	T1W	T2W	DWI	DCE
Pulse-sequence	TSE	SPAIR (TSE)	SE	TRIVE (TFE)
Flip angle	90°	90°	90°	12°
TE	7 ms	70 ms	75 ms	2.73 ms
TR	600 ms	2.8 s	3.5 s	5.5 ms
Slice thickness	3 or 3.5 mm	3 or 3.5 mm	2 mm	2 mm
Pixel resolution	0.45 × 0.45 mm²	0.55 × 0.55 mm²	1.17 × 1.17 mm²	0.75 × 0.75 mm²
Fat suppression	no	yes	yes	yes
b (s/mm^2)	/	/	700	/
Contrast agent	no	no	no	yes

508

509 **Table 2. Patient information and tumor characteristics**

Patients	**75**
Characteristics	(n=89)
BRCA1 **Mutation**	11
Mean age in years (range)	48 (26 – 75)
Tumor size in mm (range)	38 (20 – 92)
- **T1 (< 2cm)**	3
- **T2 (2-5cm)**	55
- **T3 (> 5cm and < 5cm)**	9
- **T4 (Any size tumor with direct extension to chest wall or skin)**	22
Masses	83
- **Shape:**	
o **Oval**	4
o **Round**	6
o **Irregular**	73
- **Margin:**	
o **Circumscribed**	7
o **Not circumscribed irregular**	58
o **Not circumscribed spiculated**	18
Internal enhancement characteristics:	
o **Homogeneous**	3
o **Heterogeneous**	53
o **Rim enhancement**	27
- **Kinetic curve assessment:**	
o **Persistent : type 1**	1
o **Plateau : type 2**	33
o **Washout : type 3**	39
o **Not reported**	16
- **T2 signal intensity:**	
o **Hypointense/isointense**	25
o **Hyperintense**	64
Non Mass enhancement	6
- **Distribution:**	
o **Focal**	0
o **Linear**	0
o **Segmental**	2
o **Regional**	2
o **Multiple regions**	0
o **Diffuse**	2
- **Internal enhancement patterns :**	
o **Homogeneous**	1
o **Heterogeneous**	3
o **Clumped**	2
o **Clustered ring**	0
Adenopathy	
- **Axillary**	53
- **Internal mammary**	20
Pathologic complete response	
- **Yes**	57
- **No**	32
Histologic types	
- **Infiltrating ductal carcinoma**	78
- **Infiltrating ductal carcinoma + ductal carcinoma in situ**	6
Others (metaplastic carcinoma, myoepithelial carcinoma, medullary carcinoma)	5

Type of surgery	
- **Lumpectomy**	33%
- **Mastectomy**	67%

512 **Table 3. Number of lesions imaged with DCE-MRI, T1W, T2W, and DWI sequences.**

	SUB3 (N=89)	**T1W (N=86)**	**T2W (N=76)**	**DWI (N=72)**
pCR	57	56	47	45
non-pCR	32	30	29	27

513
514

Table 4: The thirty most relevant feature classified by the ReliefF method

Order (ReliefF)	Type of tissue	Type of imaging	Type of features	Function	Parameter names
1	Tumor	T2W	Textural	SURF	std_blobs_strength
2	Parenchyma	Binary mask	Shape	Affine moment	Affine_moment_invariant_6
3	Tumor	T2W	Textural	Quadtree	sum_blocks
4	Parenchyma	T1W	Textural	GLRLM	Short_Run_Low_Gray_Level_Emphasis
5	Tumor	T2W	Textural	Quadtree	mean_c
6	Tumor	T1W	Textural	SURF	std_blobs_strength
7	Tumor	Binary mask	Shape	Zernike moment	Zernike_moment_Phi8
8	Tumor	Binary mask	Shape	Hu moment	Hu_moment_1
9	Tumor	T1W	Textural	Gabor	Gabor_square_energy_s1
10	Tumor	T2W	Textural	Quadtree	mean_r
11	Tumor	T1W	Textural	Grad	G2_y_sum
12	Tumor	DWI	Textural	SURF	mean_blobs_strength
13	Tumor	T1W	Textural	Gabor	Gabor_square_energy_s2
14	Tumor	T1W	Textural	Grad	G1_y_sum
15	Tumor	T1W	Textural	Grad	L4_sum
16	Tumor	T1W	Textural	Grad	L2_sum
17	Tumor	T1W	Textural	Grad	L1_y_sum
18	Tumor	T1W	Textural	Grad	G1_x_sum
19	Tumor	T1W	Textural	Grad	G2_x_sum
20	Tumor	T1W	Textural	Grad	L3_sum
21	Parenchyma	T1W	Textural	GLCM	Information_measure_of_correlation_2
22	Tumor	T1W	Textural	Grad	L1_x_sum
23	Tumor	T1W	Textural	Gabor	Gabor_square_energy_s5
24	Tumor	Binary mask	Shape	Skelet features	std_2
25	Parenchyma	DIFF	Textural	GLCM	Correlation
26	Tumor	DIFF	Textural	FFT features	rank_F_orient1
27	Tumor	DIFF	Textural	FFT features	rank_F_orient2
28	Tumor	DIFF	Textural	FFT features	rank_F_orient3
29	Tumor	T1W	Textural	Gabor	Gabor_square_energy_s3
30	Tumor	T2W	Textural	Quadtree	length_quadtree

Supplementary information

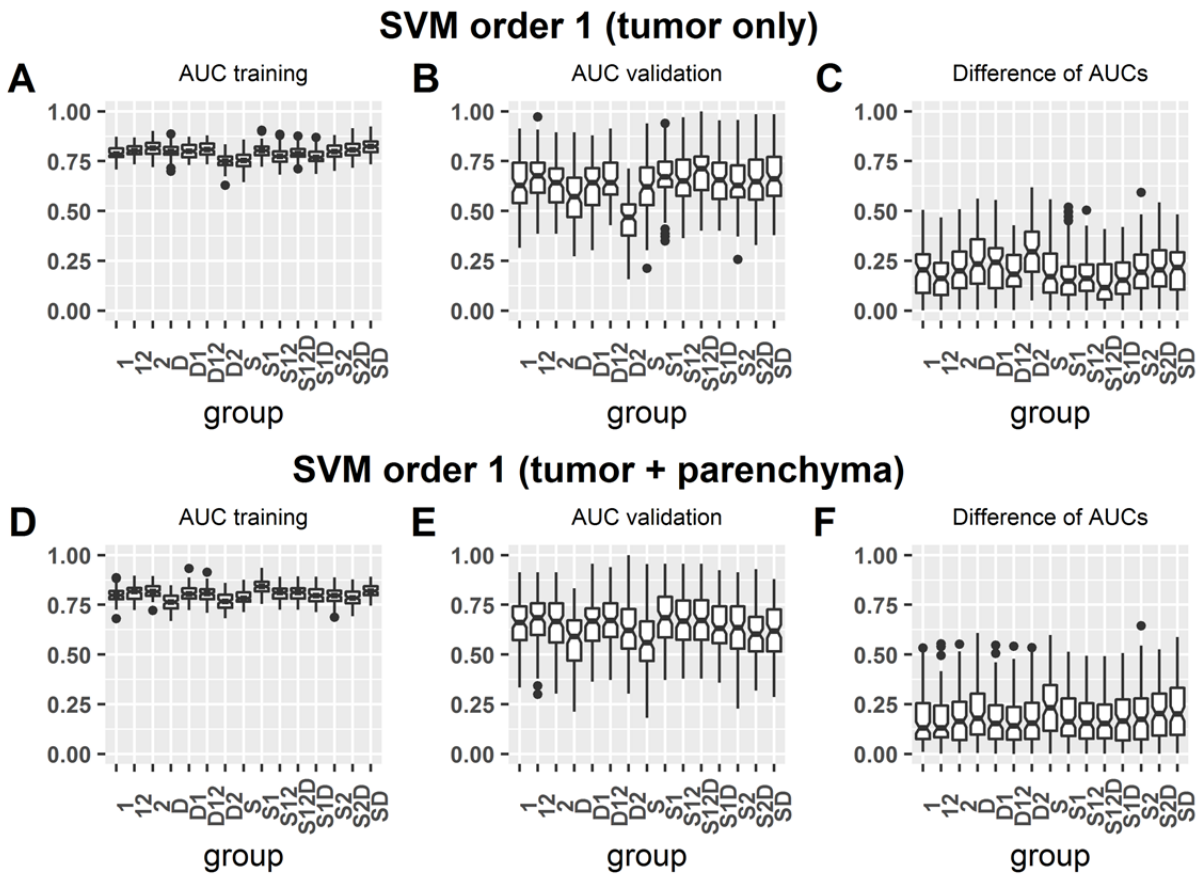
Statistical analyses

The aim of this supplementary information is to provide complete statistical analyses about the impact of different configuration of feature set on classifier performance. Indeed, in the paper, only the most relevant results were illustrated.

Multiple effects were analyzed:

1. Multiple configurations of feature sets from multi-contrast imaging (ex: combining features from the DWI and T1W).
2. Effect of adding information from parenchyma (combining features from the parenchyma and from the tumor).
3. Effect of the classifier chose.

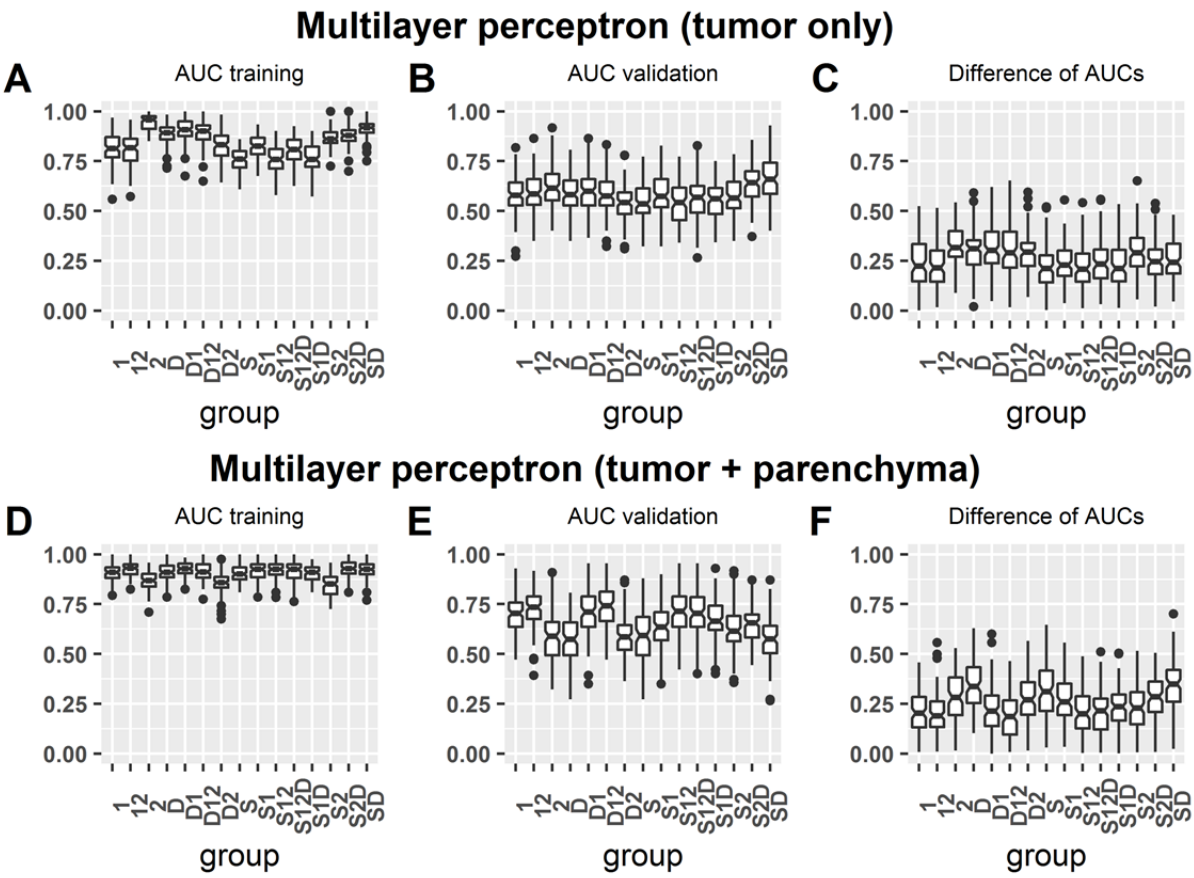
For each of the 100 random draws of training/validation set (performing with "cvpartition" Matlab function), fifteen configurations of feature set were tested. For each feature set, 4 classifiers were applied to provide a predictive model and the AUC was used to evaluate the performance of each classifier. To compare the fifteen configurations of feature, a linear mixed-effect model (LMM) was performed on AUC of validation set (or AUC of training set) for each classifier using "lmer" R function. For the design of the LMM, we considered repeated measurements on training/test set (100 different cases) with three fixed factors: Configurations (15 possibilities), Type of initial features set (tumor only or tumor + parenchyma) and Classifier (4 possibilities: SVM order 1, Random Forest, Multilayer perceptron and SVM order 2); and a random effect. The normality of each distribution was validated by a quantile-quantile plot. Figures S1, S2, S3 and S4 shows boxplot of the fifteen possible configurations with data "tumor only" for A,B and C graphs and with data "tumor + parenchyma" for D, E and F. The fixed-effect results of LMM were summarized in table S5. The LMM demonstrated a significant effect of adding information from the parenchyma with greater values of AUC_training and AUC_validation (mean difference of respectively 0.020 and 0.047 after the correction of "configurations" effect and "classifier" effect). Most of configurations using multi-contrast imaging have higher AUC values than using only one contrast; expect for the configuration "DWI + T2W". The classifier with the highest AUC values in both training and validation sets was the SVM with quadratic kernel.



551

552 **Figure S1 (Support vector machine with linear kernel): Boxplot of AUC distribution in**
553 **the training/validation set and the difference of the AUC of the two sets. Feature set used**
554 **for the classification was designed as follow: "1" designs T1-weighted imaging, "2" for**
555 **T2-weighted imaging, "D" for diffusion weighted imaging and "S" for the subtraction of**
556 **DCE-MRI. The initial feature set was composed of tumor features for A, B, C. The**
557 **initial feature set was composed of tumor and parenchyma features for D, E, and F.**

558



559

560 **Figure S2 (Multilayer perceptron classifier): Boxplot of AUC distribution in the**
561 **training/validation set and the difference of the AUC of the two sets. Feature set used for**
562 **the classification was designed as follow: "1" designs T1-weighted imaging, "2" for T2-**
563 **weighted imaging, "D" for diffusion weighted imaging and "S" for the subtraction of**
564 **DCE-MRI. The initial feature set was composed of tumor features for A, B, C. The**
565 **initial feature set was composed of tumor and parenchyma features for D, E, and F.**

566

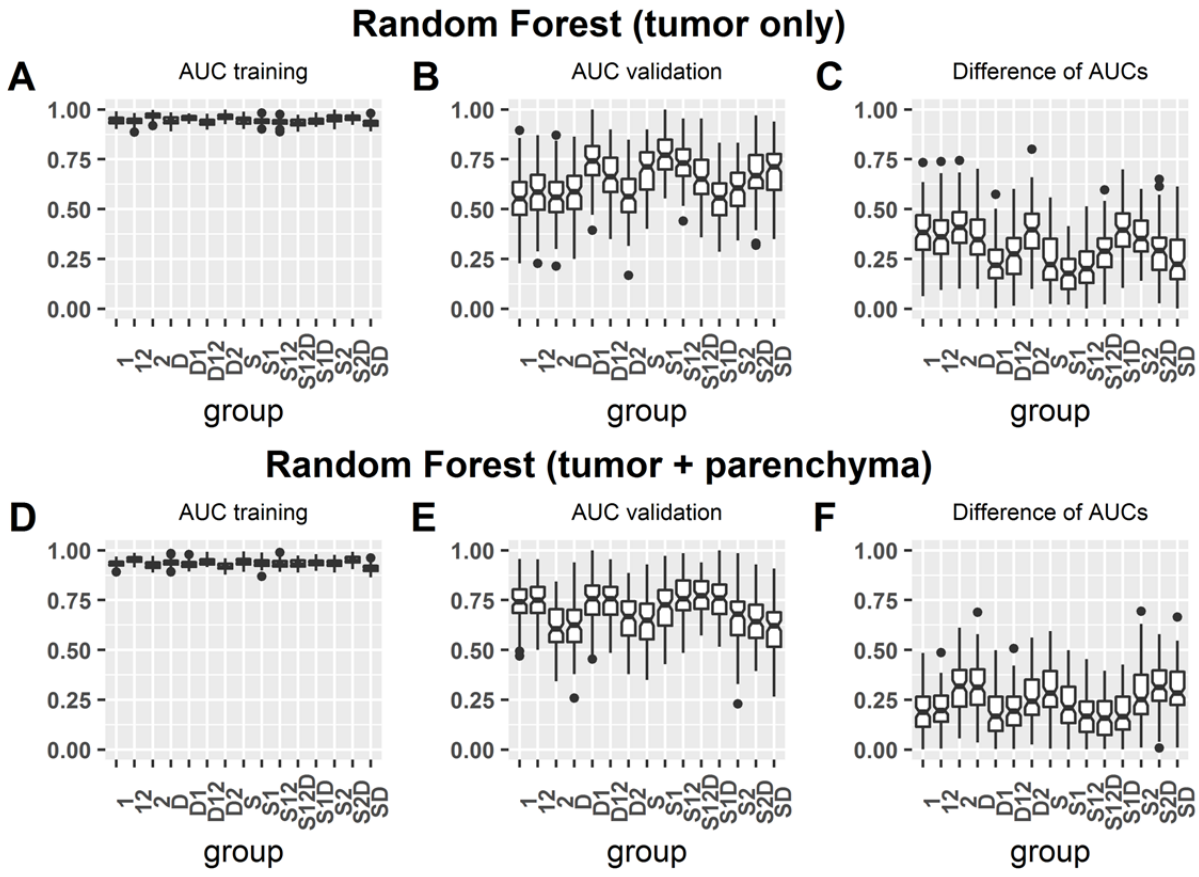
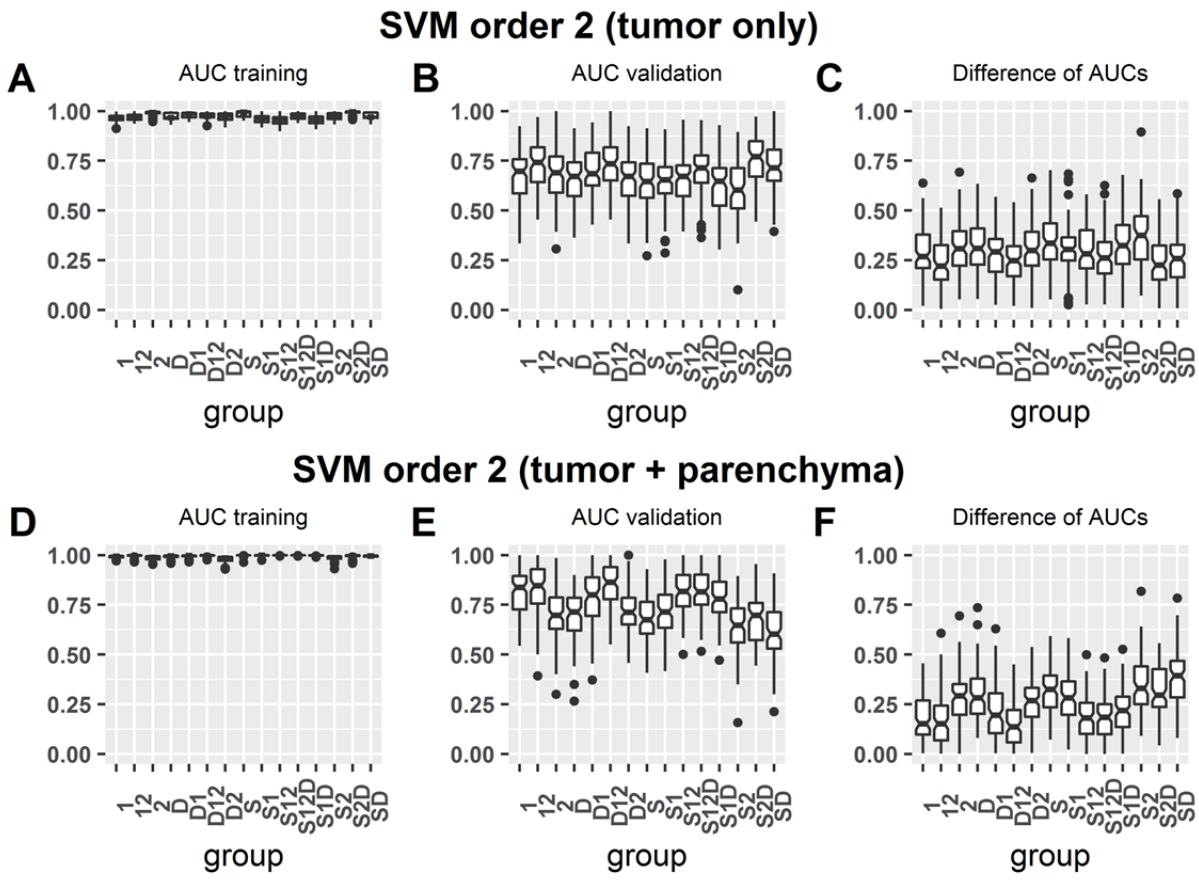


Figure S3 (Random Forest classifier): Boxplot of AUC distribution in the training/validation set and the difference of the AUC of the two sets. Feature set used for the classification was designed as follow: "1" designs T1-weighted imaging, "2" for T2-weighted imaging, "D" for diffusion weighted imaging and "S" for the subtraction of DCE-MRI. The initial feature set was composed of tumor features for A, B, C. The initial feature set was composed of tumor and parenchyma features for D, E, and F.



575

576 **Figure S4 (Support vector machine with quadratic kernel): Boxplot of AUC distribution**
577 **in the training/validation set and the difference of the AUC of the two sets. Feature set**
578 **used for the classification was designed as follow: "1" designs T1-weighted imaging, "2"**
579 **for T2-weighted imaging, "D" for diffusion weighted imaging and "S" for the**
580 **subtraction of DCE-MRI. The initial feature set was composed of tumor features for A,**
581 **B, C. The initial feature set was composed of tumor and parenchyma features for D, E,**
582 **and F.**

583

Table S5: Mean difference from reference of fixed effect of linear mixed model with random intercept applied on AUC_training results and AUC_validation results. Dunnett post-hoc analyses were made to evaluate the significance of difference. *** p<0.001

	Mean difference from reference	
	AUC Training	AUC Validation
Initial feature set		
Tumor only (reference)	--	--
Tumor + parenchyma	0.020 ***	0.047 ***
Configurations		
Sub3 (reference)	--	--
T1W	0.008 ***	0.049 ***
T1W + T2W	0.016 ***	0.073 ***
T2W	0.030 ***	0.011 (p = 0.502)
DWI	0.016 ***	-0.013 (p = 0.252)
DWI + T1W	0.028 ***	0.075 ***
DWI + T1W + T2W	0.024 ***	0.083 ***
DWI + T2W	-0.006 (p = 0.088)	-0.021 (p = 0.005)
Sub3 +T1W	0.018 ***	0.060 ***
Sub3 +T1W + T2W	0.001 (p = 1.000)	0.072 ***
Sub3 +T1W + T2W + DWI	0.010 ***	0.077 ***
Sub3 +T1W + DWI	-0.001 (p = 1.000)	0.036 ***
Sub3 +T2W	0.010 ***	0.003 (p = 1.000)
Sub3 +T2W + DWI	0.028 ***	0.040 ***
Sub3 + DWI	0.026 ***	0.022 (p = 0.002)
Classifier		
SVM order 1 (reference)	--	--
MLP	0.075 ***	-0.019 ***
Random Forest	0.143 ***	0.031 ***
SVM order 2	0.186 ***	0.073 ***