Translational Oncology 20 (2022) 101411

Contents lists available at ScienceDirect

Translational Oncology

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

Original Research

Impact of deformable registration methods for prediction of recurrence free survival response to neoadjuvant chemotherapy in breast cancer: Results from the ISPY 1/ACRIN 6657 trial

Snekha Thakran [a], Eric Cohen [a], Nariman Jahani [a], Susan P. Weinstein [a], Lauren Pantalone [a], Nola Hylton [b], David Newitt [b], Angela DeMichele [c], Christos Davatzikos [a], Despina Kontos [a,*]

[a] Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104 United States
[b] Department of Radiology and Biomedical Imaging, University of California San Francisco, San Francisco, CA 94115 United States
[c] Department of Medicine, Division of Hematology-Oncology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104 United States

ARTICLE INFO

Keywords:
Breast cancer
Deformable image registration
Longitudinal breast MRI
Parametric response maps
Radiomic features
Tumor heterogeneity
Recurrence-free survival analysis

ABSTRACT

Purpose: Image registration plays a vital role in spatially aligning multiple MRI scans for better longitudinal assessment of tumor morphological features. The objective was to evaluate the effect of registration accuracy of six established deformable registration methods(ANTs, DRAMMS, ART, NiftyReg, SSD-FFD, and NMI-FFD) on the predictive value of extracted radiomic features when modeling recurrence-free-survival(RFS) for women after neoadjuvant chemotherapy(NAC) for locally advanced breast cancer.

Methods: 130 women had DCE-MRI scans available from the first two visits in the ISPY1/ACRIN-6657 cohort. We calculated the transformation field from each of the different deformable registration methods, and used it to compute voxel-wise parametric-response-maps(PRM) for established four kinetic features.104-radiomic features were computed from each PRM map to characterize intra-tumor heterogeneity. We evaluated performance for RFS using Cox-regression, C-statistic, and Kaplan-Meier(KM) plots.

Results: A baseline model(**F1**:Age, Race, and Hormone-receptor-status) had a 0.54 C-statistic, and model **F2** (baseline + functional-tumor-volume at early treatment visit(FTV_2)) had 0.63. The F2+ANTs had the highest C-statistic(0.72) with the smallest landmark differences(5.40±4.40mm) as compared to other models. The KM curve for model F2 gave p=0.004 for separation between women above and below the median hazard compared to the model F1(p=0.31). A models augmented with radiomic features, also achieved significant KM curve separation(p<0.001) except the F2+ART model.

Conclusion: Incorporating image registration in quantifying changes in tumor heterogeneity during NAC can improve prediction of RFS. Radiomic features of PRM maps derived from warping the DCE-MRI kinetic maps using ANTs registration method further improved the early prediction of RFS as compared to other methods.

Introduction

Breast cancer is the second leading cancer in the United States and continues to increase in incidence [1]. An estimated 281,550 new cases of invasive breast cancer will be diagnosed among women in 2021 that will result in approximately 43,600 breast cancer deaths in women in the same year [1]. This emphasizes the need for early tumor diagnosis and accurate assessment of treatment response. In the budding precision medicine era, it is accepted that a predefined therapy treatment may not provide effective treatment for all patients or yield identical responses. Identifying the non-responders at an early stage may direct these patients to different therapies, leading to potentially better outcomes [2,3]. This highlights the importance of early predictions of patient response to cancer treatments or therapies [4,5]. The longitudinal patterns of tumor response may be a significant predictor in evaluating treatment response and measuring the likelihood of overall survival during neoadjuvant chemotherapy (NAC). At present, NAC has become a standard preoperational treatment paradigm for locally advanced breast cancer patients, which enables monitoring the longitudinal changes [6,7].

* Corresponding author at: Rm D702 Richards Bldg. 3700 Hamilton Walk.
E-mail address: despina.kontos@pennmedicine.upenn.edu (D. Kontos).

https://doi.org/10.1016/j.tranon.2022.101411
Received 24 January 2022; Received in revised form 9 March 2022; Accepted 27 March 2022

Breast cancer is a heterogeneous disease, with intra-tumor heterogeneity manifesting in imaging modalities both spatially and temporally. Recent studies have shown that analysis of intra-tumor heterogeneity manifests in raw image data [8,9] or parametric maps of kinetic features [10,11], which can help in the evaluation of breast cancer therapy response. Regional analysis like voxel-based analysis can capture heterogeneity better than average-based whole-tumor analysis [12,13]. Conventional biomarkers have fallen short of fully capturing intra-tumor heterogeneity as reported in previous studies [14–16]. Several studies have shown that voxel-level heterogeneity measures derived from dynamic-contrast-enhanced MRI (DCE-MRI) can be associated with NAC outcomes [11,17].

During treatment of locally advanced breast cancer, monitoring tumor changes during NAC effectively and predicting the long-term pathologic response and recurrence-free survival (RFS) are important for patient management [18]. For this purpose, clinicians refer to longitudinal breast MR images acquired in multiple visits, days or weeks apart [7]. In contrast, to quantify heterogeneous tumor changes, the challenge is to evaluate patients over multiple longitudinal MRI acquisitions to assess the morphological and functional tumor changes during treatment. For a better assessment of tissue morphological features, inter-visit image registration can play a vital role in bringing multiple MRI scans in the same space [19]. Deformable image registration techniques can also evaluate the spatial heterogeneity of voxel-level changes [18,20,21]. There are several image registration methods that exist and can be used for this task. Moreover, a comprehensive evaluation of these image registration methods remains an unexplored topic, motivating our work in this article.

The purpose of our study is to evaluate the effect of the registration accuracy of several deformable image registration methods on the predictive value of extracted radiomic features to model RFS in women undergoing NAC for locally advanced breast cancer. We also evaluate and compare six deformable image registration methods based on 2,380 landmarks individually marked by two radiologists in an independent dataset of longitudinal breast MR scans.

Methods

Study population and MRI data acquisitions

There was no consent required for this retrospective study as de-identified data was obtained from "The Cancer Imaging Archive" (TCIA) [22]. This study has been approved by the Institutional Review Board of the University of Pennsylvania. The patient population analyzed in our study was a subset of patients with longitudinal breast DCE-MRI scans, publicly available at TCIA, from the ISPY1/ACRIN 6657[7] trial. The women enrolled in the study had T3 breast tumors measuring 3cm and received anthracycline-cyclophosphamide neo-adjuvant chemotherapy. The imaging protocols were previously described according to the ISPY1/ACRIN 6657[7].

All MRI experiments were performed at 1.5 Tesla using a dedicated breast coil. Four longitudinal MRI examinations were performed at different time points during the therapy. In this study, we used MRIs from the first two visits, focusing on the ability to predict likelihood of survival early in the course of the therapy. The first MRI (T1, MRI_1 (pre-treatment)) was performed at four weeks before anthracycline-cyclophosphamide (AC) chemotherapy. The second MRI (T2, MRI_2 (early-treatment)) was performed at least two weeks after the first cycle of AC and before the second cycle of AC.

Cohort for evaluating registration accuracy

Prior to analyzing the cohort for RFS, an independent "test-retest" sample of 14 subjects with retrospectively collected breast MR images, also publicly available by the ISPY/ACRIN 6657, were used to evaluate the registration accuracy in this study [23,24]. This dataset was previously used and described in the literature to evaluate the registration accuracy for different registration methods (DRAMMS, NMI-FFD and SSD-FFD) [19]. Briefly, two experienced breast imaging radiologists annotated landmarks individually in the images. A total of 2,380 landmarks (median 78 (IQR:63.25-107.25) per patient) were marked and labeled at breast boundaries, nipples, internal milk ducts, chest walls, glandular structures, vessels, and tumors. We registered the follow-up image to the baseline image for each patient (median 17.5 (IQR:14-55.75) days apart). The expert-defined landmarks were used as the gold standard. In the current study, we included and compared the registration accuracy of ANTs, ART and NiftyReg methods with above mentioned methods.

Cohort for the prediction of NAC RFS

The functional tumor volume (FTV) calculated at pre-treatment (FTV_1) and early-treatment visits (FTV_2) was available for each DCE-MRI scan. According to STEEP criteria [25], RFS outcomes were calculated for each woman, defined as the time from the first chemotherapy cycle to the event (i.e., recurrence or death). The inclusion and exclusion criteria used to attain the final study cohort are summarized in Figure 1. After excluding 65 scans due to incomplete clinical and/or imaging data at the first two visits, 15 scans due to the low image quality, and 12 scans for which image registration failed, our final analysis included 130 women with 38 events from the ISPY1/ACRIN 6657[7] cohort for RFS analysis. Pathologic Complete Response (pCR) information was missing for 5 participants, leaving 125 patients for pCR analysis, having an overall tumor response rate for our study sample of about 28.5%.

Image pre-processing

All images were pre-processed using N3 bias-field normalization to minimize intensity variations caused by the imaging sequences or artifacts [26] . Histogram matching was applied between pre- and early-treatment images for more accurate registration [10]. These image pre-processing steps were applied before registration and quantitative analysis.

Registration methods and parameter setting

After the pre-processing steps, we used six different deformable image registration methods outlined in Supplementary Table S1 (ANTs, DRAMMS, ART, NiftyReg, SSD-FFD, and NMI-FFD) for spatial and anatomical alignment of the early-treatment MRI scans to their corresponding pre-treatment MRIs. Each deformable image registration method is usually described by three important components: 1)

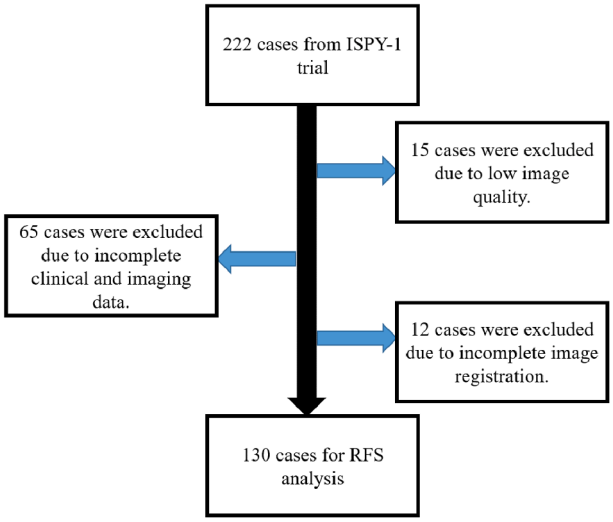

Fig. 1. Inclusion and exclusion criteria for our study.

deformation model, 2) similarity metric, and 3) regularization, as shown in Table S1. We chose these methods because they are publicly available and previously reported in the literature to register MR images in general [19,27–31]. Below we summarize the registration methods evaluated in our study. The parameters stated below were not set arbitrarily. Generally, the authors of the below-mentioned studies evaluated a wide variety of parameter settings and reported the ones that gave the most accurate alignment of longitudinal breast MR images.

Advanced normalization tools (ANTs)

ANTs [32] splits the diffeomorphic deformation into two symmetric components. Instead of deforming the first image into the space of the second image, which is usually not symmetric to the input images, ANTs work by deforming two input images, each towards the "midpoint" image. As a result, the formulation is symmetric to the input images. We used cross-correlation with symmetric image normalization method (SyN) transformation in this study. The ANTs registration tool can be found at http://stnava.github.io/ANTs/.

Deformable registration via attribute matching and mutual-saliency weighting (DRAMMS)

This deformable image registration method [33] is based on attribute-matching and mutual-saliency-weighting, which uses high-dimensional Gabor texture attributes obtained from the multi-scale and multi-orientation neighborhoods of voxels [31]. The DRAMMS registration tool is publicly available at http://www.cbica.upenn.edu/sbia/software/dramms/. The regularization weight (g) is an important parameter that may need to be tuned in the DRAMMS registration method. It is commonly set between 0 and 1. We used g=0.3, which led to slightly smoother deformations than the default setting (g=0.2).

Automatic registration toolbox (ART)

In the ART [28,34] method, the intensity values of all the neighborhood voxels are stacked to generate a high dimensional feature vector that characterizes a voxel. The inner-product of two feature vectors and an idempotent and symmetric matrix that removes the mean of the vector it pre-multiplies defines the similarity of two voxels. ART only searches correspondences at voxels whose gradient norms are in a certain upper percentile of the gradient magnitude histogram when analyzing the highest image resolution. For ART, sd is defined as specifying the degree of smoothing applied to the displacement vector field, which typically ranges from 5.0 to 12.0 mm. we set sd=8.0 in our evaluation. This ART registration tool is publicly available at http://www.nitrc.org/projects/art/.

Free form deformation (FFD)

FFD is a geometric transformation model and combined with normalized mutual information (NMI) similarity metric in the original work [30]. It can be used along with other similarity metrics. Two important parameters are control point spacing δ with a larger spacing that capture more global deformations; and regularization weight λ with larger weights that provide smoother deformations. We used δ = 10 mm, larger spacing means capturing more global deformations; and λ =0.01 in our evaluation. NMI-FFD and SSD-FFD were included in this study. This software tool is publicly available at http://www.doc.ic.ac.uk/~dr/software/.

Nifty registration (NiftyReg)

This method is used for global and local image registration [27,30]. The global registration uses the block-matching technique whereas cubic B-splines to capture the local registration. This method is based on free form deformation [30], further extended as fast free form deformation [27] to speed-up registration. Normalized Mutual Information (NMI) and the Bending-Energy are used in this study. This software tool is publicly available at https://www.nitrc.org/projects/niftyreg/.

Evaluation criterion for registration accuracy

The inter-expert landmark differences and algorithm-to-expert landmark differences were first computed to evaluate the registration accuracy using the previously described method [19]. Two radiologists independently found their corresponding points in the follow-up image, denoted as $y1_i$, and $y2_i$. The inter-expert landmark difference (the length of the solid line as shown in Figure 2), for each landmark, was computed as the Euclidean distance between the landmark points designated by each of the two experts.

$$d\ (y1_i,\ y2_i),\ \text{for}\ i=1,2,3,\ldots\ldots\ldots\ldots\ldots\ldots,k \quad (1)$$

The algorithm-to-expert landmark difference (the average length of the dashed lines in Figure 2) was computed as:

$$[d\ (y3_i,\ y1_i) + d\ (y3_i,\ y2_i)]/2,\ \text{for}\quad i=1,2,3,\ldots\ldots\ldots\ldots,k \quad (2)$$

where d (•,•) is again the Euclidean distance and i varies over the k landmarks.

Computation of parametric response maps (PRM) and their radiomic features for the prediction of RFS

DCE-MRI images consisted of a pre-contrast (t_0) image and images at two times, t_1 and t_2, following injection of the contrast agent. The signal intensity of each voxel was computed at each time point $S(t)$. Four voxel-wise kinetic features were calculated from the t_0, t_1, and t_2 intensity values at each voxel, for both pre- and early-treatment images: peak enhancement (PE), signal enhancement ratio (SER), wash-in slope (WIS), and wash-out slope (WOS) [18]. We also computed the difference between the kinetic feature maps from pre- and early-treatment MR images without registration. Resampling value of 1 mm was used for higher accuracy. Rotation variant LBP was used and a total of 27 neighbors were considered for computation of LBP. Linear Interpolation was used for the MRI images. We extracted 104 radiomic texture features (detailed list of radiomic features in Supplementary Table S2) from each without registered kinetic features map, using the publicly available software, Cancer Phenomics Toolkit (CaPTk) [35], applied principal component (PC) analysis to the 104-dimensional feature vector after calculating the z-score, and retained the first four PCs for modeling (i.e., one covariate for every 10 events in our cohort). To quantify voxel scale changes between visits, corresponding voxels between pre- and early-treatment visits maps must be determined. The transformation field derived from the image registration allowed us to construct the voxel-wise maps of change in a given kinetic feature (also called the PRM) between corresponding points at the pre- and early treatment visits. For each of the six different deformable registration methods (ANTs, DRAMMS, ART, NiftyReg, NMI-FFD, and SSD-FFD), we calculated the transformation field and used it to warp the kinetic maps obtained from early-treatment MR images. For a voxel **x** in the pre-treatment image, let $T_{12}(\mathbf{x})$ indicate the voxels in the early-treatment image, which was warped to **x**. For a given Kinetic

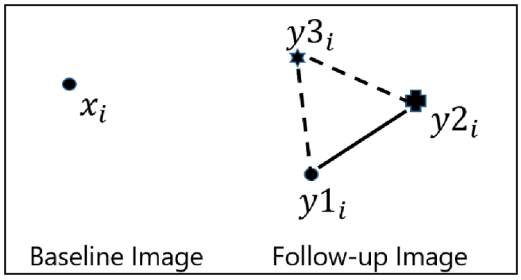

Fig. 2. Depiction of inter-expert and algorithm versus expert landmark differences.

Feature (KF), one of PE, SER, WIS, or WOS, we computed at the pre- or early-treatment visit(KF_1 or KF_2), the PRM value at each voxel $\mathbf{x}$[18] is calculated as:

$$PRM(x)=Jacobian\ (KF_2(T_{12}(\mathbf{x}))\ -\ KF_1(x)) \quad (3)$$

Jacobian is the proportional volume change at **x** between pre- and early-treatment visits, which scales the value when a voxel in one image is the larger or smaller volume in the corresponding image. The kinetic features (and their PRMs and radiomic features) were only measured within tumor. A signal enhancement ratio method was used to analyze DCE-MR images and segment functional tumor volumes (FTV_1 and FTV_2). A total of 104 radiomic features were extracted from each PRM kinetic map, using CaPTk [35], applied principal component (PC) analysis to the 104-dimensional feature vector after calculating the z-score, and retained the first four PCs for modeling (i.e., one covariate for every 10 events in our cohort). Figure 3 shows the flow chart of the analysis performed in this study.

Statistical analysis

Cox proportional hazards regression was used to model 5-year RFS for each set of covariates, comparing eight models: 1) First, a model with the baseline covariates of age, race, and hormone receptor status (model F1), 2) model F1 plus functional tumor volume at the early-treatment visit (FTV_2) (model F2), 3) model F2 plus radiomic feature without registration(model F3) and 4-9) the covariates in F2 with the addition of the radiomic feature PCs derived from each registration method (ANTs, DRAMMS, ART, NiftyReg, SSD-FFD, and NMI-FFD). The C-statistic [36] was calculated both on the full models as well as 3-fold cross-validation averaged over 100 replicates. The full models were used to generate corresponding KM plots, and the log-rank *p*-value was used for estimating significance of KM separation and hazard ratios (HR) in Cox modeling. Each subject's risk signature was dichotomized at the median into high- and low-risk groups. For a given model, the risk signature of each subject was defined as that subject's values of the covariates in the model (age, race, hormone receptor status, FTV_2, and selected features) weighted by the corresponding coefficients of those covariates in the model, to arrive at a predicted risk score. For every model, we also used a randomization test to calculate a *p*-value versus the null hypothesis that the improvement in the C statistic, obtained by adding the radiomic data, was due only to chance. This randomization test was performed for both full-model and 3-cross-validated with 100 repetitions C-statistic. The *p*-value of 0.05 cutoff was used to indicate statistical significance throughout this article. The radiomic features were extracted using Cancer Imaging Phenomics Toolkit (CaPTK) [35] version 1.7.2. The statistical analysis was computed using R version 3.3.2.

Results

Landmark-based registration accuracy

The mean and SD of landmark difference(mm) are listed for different registration methods (Table 1). The mean inter-expert landmark differences was 3.12 ± 2.84 mm. The landmark differences of all the automated registration methods evaluated were larger than the inter-expert differences. Among the automated registration methods included in our study, ANTs had the smallest landmark differences (5.40

Table 1
Mean and Standard Deviation (SD) of Landmark differences(mm) for different registration methods and inter-expert.

	Inter-expert	ANTs	DRAMMS	ART	NiftyReg	NMI-FFD	SSD-FFD
Mean	3.12	5.40	6.05	10.33	7.02	8.21	9.46
SD	2.84	4.40	4.86	2.79	4.23	3.81	4.55

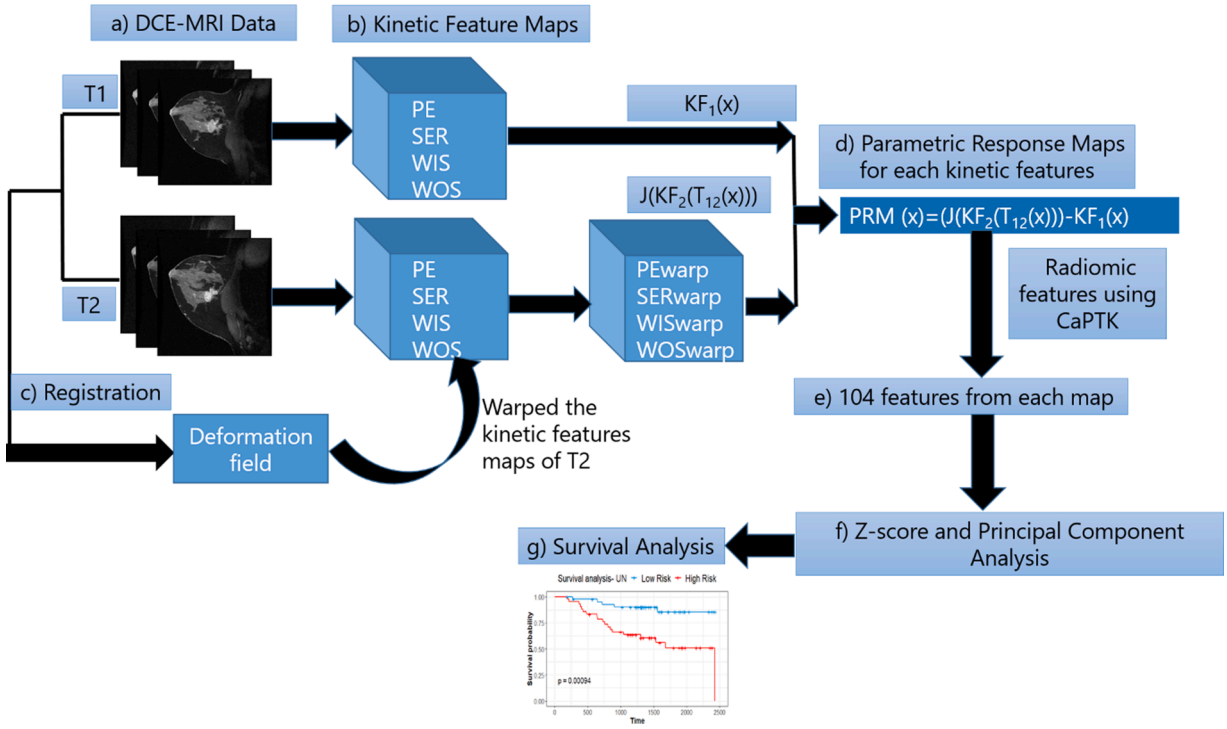

Fig. 3. Outline of the process using different deformable image registration methods. (**a**) Acquiring dynamic contrast-enhanced magnetic resonance imaging before and during neoadjuvant chemotherapy. (**b**) Computing different kinetic maps from both pre-treatment and early treatment MR images. (**c**) Applying different deformable image registration method to obtain the transformation field, followed by warping the kinetic maps obtained from early treatment MR images. (d) Parametric response maps (PRM) showing kinetic feature variation. (**e**) Computing radiomic features extraction followed by principal component analysis. (f) Building multi-variable models using the meta feature and comparing their performance in predicting recurrence-free survival (RFS).

± 4.40 mm) and was closest to the inter-expert differences, followed by the DRAMMS (6.05 ± 4.86 mm), NiftyReg (7.02±4.23), NMI-FFD (8.21 ± 3.81 mm), and SSD-FFD (9.46 ± 4.55 mm). ART had the largest landmark differences (10.33 ± 2.79 mm) from the inter-expert differences.

Recurrence-free survival

130 women were used in our study for RFS analysis. The clinical, demographic, and histopathologic information were available for each patient, as shown in Table 2. In this study cohort, 38 women (29.2%) had recurrence/death events, and 92 women (70.8%) were event-free until their last available follow-up (Table 2).

The baseline model (F1: age, race, and hormone receptor status) provided a C-statistic of 0.54 with a full-model and mean 3-fold cross-validated with 100 repetitions C-statistic of 0.44 (SD:0.07) (Table 3). Fitting this baseline (F1) model to the full dataset, none of the covariates (age, race, and hormone receptor status) showed a statistically significant association with RFS (Table 4). When adding FTV_2 to the baseline model (F2), this F2 model provided a full-model C-statistic of 0.63 and a mean 3-fold cross-validated with 100 repetitions C-statistic of 0.53 (SD:0.08) (Table 3). Fitting this model to the full dataset, only FTV_2 showed a statistically significant association with RFS (hazard ratio: 1.62, $p < 0.001$), as shown in Table 5.

For both full and 3-fold cross-validated with 100 repetitions models, the C-statistic of Cox models including F2 + PCs of PRM radiomic features, as derived from different registration methods were higher than the C-statistic of the baseline (F1) model, F2 model and the model incorporating radiomic features without registration. The F2 + radiomics without registration model (F3 model) provided a C-statistic of 0.66 and a mean 3-fold cross-validated with 100 repetitions C-statistic of 0.55 (SD:0.07). Among the automated registration methods (Table 6), F4(F2+ANTs) model had the highest performance for the full model (C-statistic of 0.72 with $p = 0.01$) as well as a mean 3-fold cross-validated with 100 repetitions (C-statistic of 0.63(SD: 0.07) with $p = 0.01$). F5 (F2+DRAMMS) model provided a C-statistic of 0.70 with a full-model and mean 3-fold cross-validated with 100 repetitions C-statistic of 0.58(SD:0.07), followed by F2+NiftyReg, F2+NMI-FFD, and F2+SSD-FFD. F6 (F2 + ART) had the lowest performance for both full (C-statistic of 0.66) and 3-fold cross-validation with 100 repetitions (mean C-statistic of 0.54(SD: 0.08)).

When combining the FTV_2 with the baseline predictors, the KM plots (Figure 4) showed a greater separation between high-risk and low-risk patients based on their median risk score with log-rank $p = 0.004$ as compared to the baseline predictors (log-rank $p = 0.31$).

Table 2
Patient characteristics for our study from the ISPY-1/ACRIN 6657 cohort.

	Non-recurrent cases N=92(70.8%)	Recurrent/death cases N=38(29.2%)	*P*-Value
Age			
Median age (IQR), years	47.73(41.31-53.30)	48.77(39.16-53.69)	
Laterality			$p = 0.78$
Left	46(50%)	18(47.4%)	
Right	46(50%)	20(52.6%)	
Race			$p = 0.63$
White or Hispanic	69(75%)	30(78.9%)	
Others	23(25%)	8(21.1%)	
Progesterone Receptor Status			$p = 0.89$
Negative	52(56.5%)	21(55.3%)	
Positive	40(43.5%)	17(44.7%)	
Hormone Response Status			$p = 0.38$
HER2+	30(32.6%)	17(44.7%)	
HR+ and HER2-	39(42.4%)	12(31.6%)	
Triple-negative	23(25%)	9(23.7%)	

Table 3
C-statistics of the F1: baseline and F2: baseline + FTV_2 for full and 3-fold cross-validation (with 100 rep) Cox models. SD: Standard Deviation; rep: repetitions.

Model	Covariates	C-statistic (Full model)	Mean C-statistic (SD) (3-fold cross-validation with 100 rep)
F1: Baseline	Age, Race, Hormone receptor status	0.54	0.44(0.07)
F2: F1+FTV_2	Age, Race, Hormone receptor status, FTV_2	0.63	0.53(0.08)
F3: F2+Without Registration	Age, Race, Hormone receptor status, FTV_2, radiomics features without registration.	0.66	0.55(0.07)

Table 4
Multi-variable analyses of the F1: baseline model using the full dataset for hazard ratio assessment.

Covariate	Hazard ratio	95% CI	p-value
Age	0.99	0.95-1.03	0.72
Race	0.91	0.41-1.99	0.82
Hormone receptor status	1.11	0.74-1.67	0.59

Table 5
Multi-variable analyses of the F2: baseline and FTV_2 model using the full dataset for hazard ratio assessment. * $p < 0.05$

Covariate	Hazard ratio	95% CI	p-value
Age	0.98	0.95-1.02	0.59
Race	0.85	0.39-1.88	0.69
Hormone receptor status	0.99	0.64-1.52	0.97
FTV_2	1.62	1.19-2.20	<0.001*

The F4 model gave a greater separation between high-risk and low-risk patients with log-rank $p<0.0001$ as compared to other models (F5: log-rank $p = 0.00061$, F6: log-rank $p = 0.036$, F7: log-rank $p=0.00079$, F8: log-rank $p = 0.0023$, and F9: log-rank $p = 0.0012$) as shown in Figure 5.

Discussion and conclusion

It has been shown that early-treatment response assessment helps in optimizing patient care and treatment alteration [37,38]. Voxel-wise 3-D volumetric changes can be automatically captured by deformable image registration. This makes it a potentially suitable choice for quantifying the spatially heterogeneous changes within tumors over the course of therapy. This study evaluated six different deformable registration algorithms (ANTs, DRAMMS, ART, NiftyReg, SSD-FFD, and NMI-FFD) for registration of longitudinal breast DCE-MRI scans. We evaluated the effect of the registration accuracy of these registration methods on the predictive value of radiomic features to model RFS in women undergoing NAC for locally advanced breast cancer. We also compared the accuracies of different registration methods using expert-defined landmarks as gold standard.

Our work adds several important findings to the literature. We showed preliminary results of using the ANTs registration method, which had the best accuracy in registering the longitudinal breast MR images as evidenced by the smallest landmark differences and being closest to the inter-expert differences. DRAMMS had earlier demonstrated smaller landmark difference reported in previous studies [19], which were almost 25–40% smaller than the difference from other registration methods (CC-FFD, NMI-FFD, SSD-FFD, Demons, and Diffeomorphic Demons). Our results suggest that ANTs registration outperforms other methods (DRAMMS, ART, NiftyReg, SSD-FFD, and

Table 6
Mean C-statistics of full and 3-fold cross validation (with 100 rep) Cox models including F2 + PCs of PRM radiomic features, as derived from each registration method. P value computed using randomization test. SD:Standard Deviation.* p <0.05

Model	C-statistic (Full model)	p-value	Mean C-statistic (SD) (3-fold cross-validation with 100 rep)	p-value
F4: F2+ANTs	0.72	0.01*	0.63(0.07)	0.01*
F5: F2+DRAMMS	0.70	0.05	0.58(0.07)	0.07
F6: F2+ART	0.66	0.20	0.54(0.08)	0.16
F7: F2+NiftyReg	0.68	0.06	0.56(0.08)	0.20
F8: F2+NMI-FFD	0.67	0.11	0.55(0.07)	0.07
F9: F2+SSD-FFD	0.67	0.14	0.56(0.08)	0.17

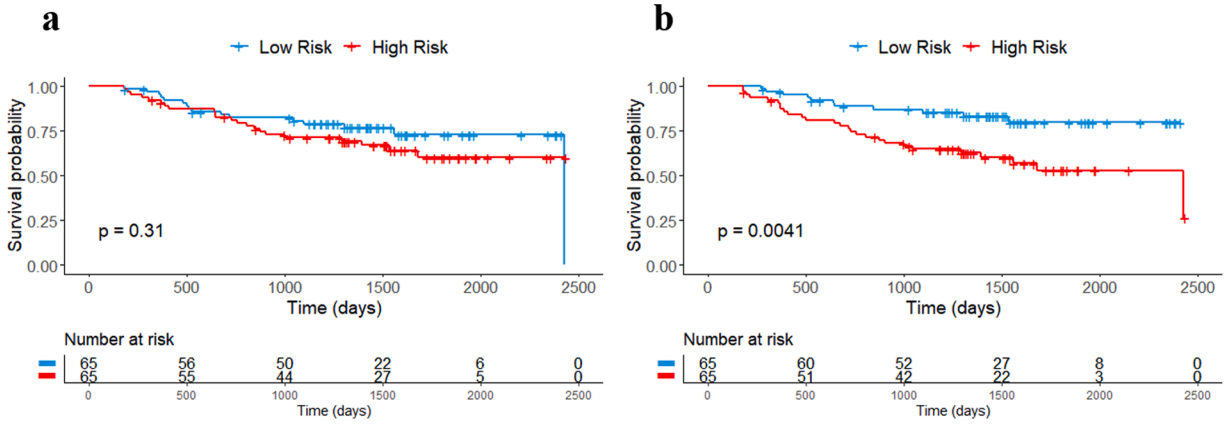

Fig. 4. Kaplan-Meier plots of RFS for full models. a) F1: Baseline, and b) F2: F1+FTV_2.

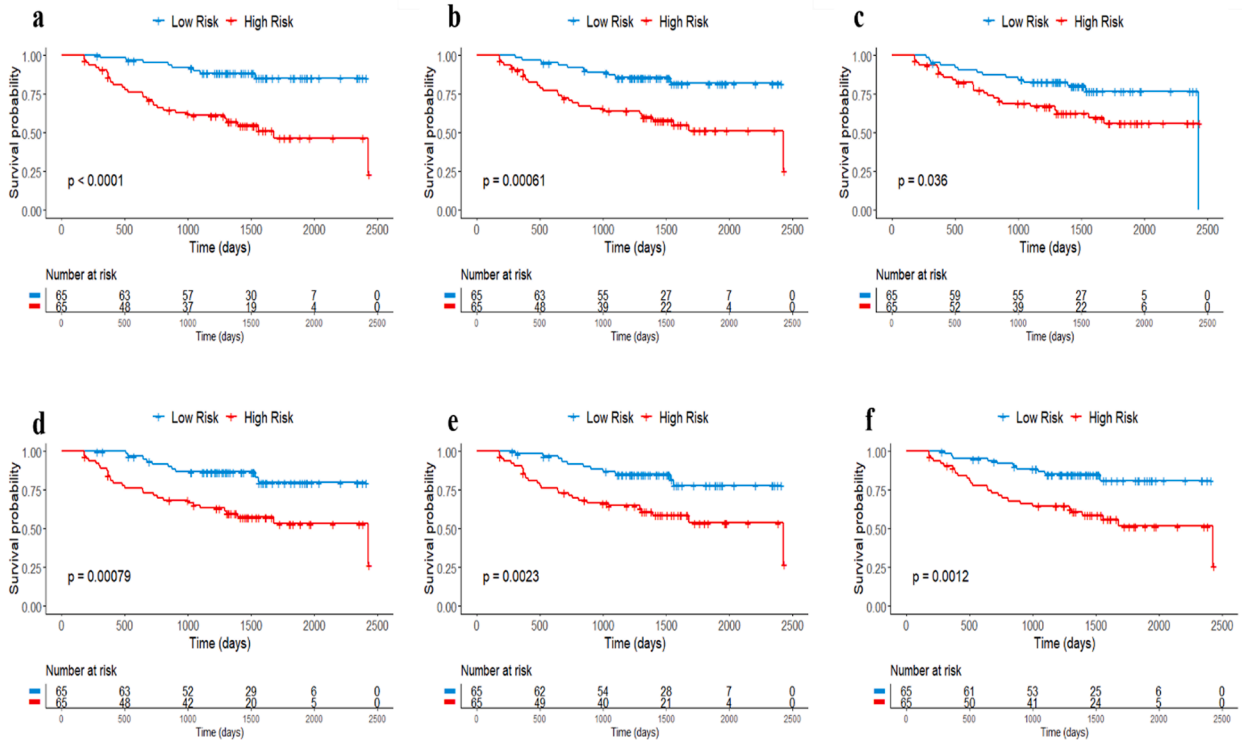

Fig. 5. Kaplan-Meier plots of RFS (full models) for different registration methods. a) F2+ANTs, b) F2+DRAMMS, c) F2+ART, d) F2+Niftyreg, e) F2+NMI-FFD, and f) F2+SSD-FFD.

NMI-FFD) in terms of registration accuracy. The registration process improves the alignment of the input images, hence ideally resulting in improved performance of extracted PRM features. Several studies reported for prediction of RFS without integration of image registration [39–41]. Our results suggest that indeed registration methods can improve early prediction of RFS, regardless of which method is used. Furthermore, improvement in registration accuracy also results in further improvement of the extracted PRM radiomic features. Our results suggest that radiomic features of PRM maps using the ANTs registration method significantly improved the early prediction of

survival during NAC compared to other methods.

In our study, we extracted PRMs derived from warping the DCE-MRI kinetic maps using registration methods. The radiomic features based on the PRMs of kinetic features are also important in capturing tumor heterogeneity to augment models of RFS. The framework of the current study consists of robust registration, voxel-wise changes of PRMs of kinetic features, radiomic feature and principal component analysis provide statistically significant improvements over previous similar analyses in predicting RFS [7,14]. By adding registration, we improved the PRM feature performance, with further improvement for registration methods that are most accurate.

Important limitations of our study must be noted. We analyzed a relatively small sample size (n=130) of the women with small number of events for RFS analysis. We plan to perform another round of evaluation when the imaging data from a larger independent validation data set such as ISPY-2 becomes available. An extension of this study can be done by applying these analyses to longitudinal images at mid- and late-treatment time points to better characterize heterogeneous tumor responses and the effects of treatment over time. The accuracy of these registration methods will also be tested for large anatomical variabilities or imaging device differences (e.g., 2-D ultrasound to 3-D MRI) in future studies. Other registration methods such as DROP, HAMMER, Plastimatch, and MIND will also be evaluated in future studies. Registration accuracy can be further improved by to utilizing multiple registration methods in a meta-analysis framework, where an underperforming or failed method in a task can be overcome by other methods. Preliminary results for the evaluation of meta-registration pipelines have been published in recent articles. For example, the combination of ANTs and DRAMMS in a multi-atlas labeling framework, results in a high-performance method, as shown in a recent MICCAI segmentation challenge [42]. Another study focused on the combination of ANTs, NiftyReg, and DROP and showed significantly reduced registration errors in pulmonary images [43]. Such a meta-registration framework can also be explored for longitudinal breast MRI in the future studies and may provide even more predictive signatures in treatment response assessment.

In conclusion, we evaluated the effect of the registration accuracy of several deformable registration methods on the predictive value of radiomic features to model RFS in women undergoing NAC for locally advanced breast cancer. Our study also performed a comprehensive evaluation of the deformable image registration methods (ANTs, DRAMMS, ART, NiftyReg, SSD-FFD, and NMI-FFD) for longitudinal breast MR images to evaluate registration accuracy, using expert-annotated landmarks as gold standard. Our results showed a two-fold conclusion that analyzing registration-quantified tumor changes improves the prediction accuracy for RFS analysis and improvement in the registration process results in further improvements in the performance of PRM radiomic features when predicting RFS. In our study, the radiomic features of PRM maps derived from warping the DCE-MRI kinetic maps using the ANTs registration method significantly improved early prediction of survival during NAC.

CRediT authorship contribution statement

Snekha Thakran: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft. **Eric Cohen:** Formal analysis, Writing – review & editing. **Nariman Jahani:** Software, Writing – review & editing. **Susan P. Weinstein:** Methodology, Writing – review & editing. **Lauren Pantalone:** Investigation, Project administration, Writing – review & editing. **Nola Hylton:** Investigation, Writing – review & editing. **David Newitt:** Investigation, Writing – review & editing. **Angela DeMichele:** Validation, Writing – review & editing. **Christos Davatzikos:** Methodology, Investigation, Writing – review & editing. **Despina Kontos:** Conceptualization, Methodology, Visualization, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the NIH grant: 5R01CA197000-05.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tranon.2022.101411.

References

[1] American Cancer Society. Cancer Society, Cancer Facts & Figures 2021, American Cancer Society, Atlanta, 2021.
[2] I.I. Wistuba, J.G. Gelovani, J.J. Jacoby, S.E. Davis, R.S Herbst, Methodological and practical challenges for personalized cancer therapies, Nat. Rev. Clin. Oncol. 8 (2011) 135–141.
[3] E. Huang, MD McNeese, EA Strom, GH Perkins, A Katz, GN Hortobagyi, V Valero, HM Kuerer, SE Singletary, KK Hunt, AU Buzdar, TA Buchholz, Locoregional treatment outcomes for inoperable anthracycline-resistant breast cancer, Int. J. Radiat. Oncol. Biol. Phys. 53 (2002) 1225–1233.
[4] M. Dowsett, A.K Dunbier, Emerging biomarkers and new understanding of traditional markers in personalized therapy for breast cancer, Clin. Cancer Res. 14 (2008) 8019–8026.
[5] U. McDermott, J. Settleman, Personalized cancer therapy with selective kinase inhibitors: an emerging paradigm in medical oncology, J. Clin. Oncol. 27 (2009) 5650–5659.
[6] A. Fangberget, L.B. Nilsen, K.H. Hole, M.M Holmen, O. Engebraaten, B. Naume, H. J. Smith, D.R. Olsen, T. Seierstad, Neoadjuvant chemotherapy in breast cancer-response evaluation and prediction of response to treatment using dynamic contrast-enhanced and diffusion-weighted MR imaging, Eur. Radiol. 21 (2011) 1188–1199.
[7] N.M. Hylton, J.D. Blume, W.K. Bernreuter, E.D. Pisano, M.A. Rosen, E.A. Morris, P. T. Weatherall, C.D. Lehman, G.M. Newstead, S. Polin, H.S. Marques, L.J. Esserman, M.D. Schnall, Locally advanced breast cancer: MR imaging for prediction of response to neoadjuvant chemotherapy – Results from ACRIN 6657/I-SPY TRIAL, Radiology 263 (2012) 663–672.
[8] J.R. Teruel, M.G. Heldahl, P.E. Goa, M. Pickles, S. Lundgren, T.F. Bathen, P. Gibbs, Dynamic contrast-enhanced MRI texture analysis for pretreatment prediction of clinical and pathological response to neoadjuvant chemotherapy in patients with locally advanced breast cancer, NMR Biomed. 27 (2014) 887–896.
[9] J. Parikh, M. Selmi, G. Charles-Edwards, J. Glendenning, B. Ganeshan, H. Verma, J. Mansi, M. Harries, A. Tutt, V. Goh, Changes in primary breast cancer heterogeneity may augment midtreatment MR imaging assessment of response to neoadjuvant chemotherapy, Radiology 272 (2014) 100–112.
[10] D.I. Golden, J.A. Lipson, M.L. Telli, J.M. Ford, D.L Rubin, Dynamic contrast-enhanced MRI-based biomarkers of therapeutic response in triple-negative breast cancer, J. Am. Med. Informatics Assoc. 20 (2013) 1059–1066.
[11] A. Ashraf, B. Gaonkar, C. Mies, A. DeMichele, M. Rosen, C. Davatzikos, D. Kontos, Breast DCE-MRI kinetic heterogeneity tumor markers: Preliminary associations with neoadjuvant chemotherapy response, Transl. Oncol. 8 (2015) 154–162.
[12] M. Mahrooghy, A.B. Ashraf, D. Daye, E.S. McDonald, M. Rosen, C. Mies, M. Feldman, D. Kontos, Pharmacokinetic tumor heterogeneity as a prognostic biomarker for classifying breast cancer recurrence risk, IEEE Trans. Biomed. Eng. 62 (2015) 1585–1594.
[13] X. Li, L.R. Arlinghaus, G.D. Ayers, A.B. Chakravarthy, R.G. Abramson, V. G. Abramson, N. Atuegwu, J. Farley, I.A. Mayer, M.C. Kelley, I.M. Meszoely, J. Means-Powell, A.M Grau, M. Sanders, S.R. Bhave, T.E. Yankeelov, DCE-MRI analysis methods for predicting the response of breast cancer to neoadjuvant chemotherapy: pilot study findings, Magn. Reson. Med. 71 (2014) 1592–1602.
[14] N.M. Hylton, C.A Gatsonis, M.A. Rosen, C.D. Lehman, D.C. Newitt, S.C. Partridge, W.K. Bernreuter, E.D. Pisano, E.A. Morris, P.T. Weatherall, S.M. Polin, G. M. Newstead, H.S. Marques, L.J. Esserman, M.D. Schnall, ACRIN 6657 Trial Team and I-SPY 1 TRIAL Investigators, Neoadjuvant chemotherapy for breast cancer : functional tumor volume by mr imaging predicts recurrence-free survival — results from the ACRIN 6657 /CALGB 150007 I-SPY 1 TRIAL, Radiology 279 (2015) 44–55.
[15] J.H. Kim, E.S. Ko, Y. Lim, K.S. Lee, B.K. Han, E.Y. Ko, S.Y. Hahn, S.J. Nam, Breast cancer heterogeneity : MR imaging texture analysis and survival outcomes, Radiology 282 (2016) 665–675.
[16] J. Wu, G. Cao, X. Sun, J. Lee, D.L. Rubin, S. Napel, A.W. Kurian, B.L. Daniel, R. Li, Intratumoral spatial heterogeneity at perfusion MR imaging predicts recurrence-free survival in locally advanced breast cancer treated with neoadjuvant chemotherapy, Radiology 288 (1) (2018) 26–35.

[17] N Bhooshan, ML Giger, SA Jansen, Li Lan, Cancerous breast lesions on dynamic contrast-enhanced MR images: computerized characterization for image-based prognostic markers, Radiology 254 (3) (2010 Mar) 680–690.
[18] N. Jahani, E. Cohen, M.K. Hsieh, S.P. Weinstein, L. Pantalone, N. Hylton, D. Newitt, C. Davatzikos, D. Kontos, Prediction of treatment response to neoadjuvant chemotherapy for breast cancer via early changes in tumor heterogeneity captured by DCE-MRI registration, Sci. Rep. 9 (1) (2019) 1–12, https://doi.org/10.1038/s41598-019-48465-x.
[19] Y. Ou, S.P. Weinstein, E.F. Conant, S. Englander, X. Da, B. Gaonkar, M.K. Hsieh, M. Rosen, A. DeMichele, C. Davatzikos, D. Kontos, Deformable registration for quantifying longitudinal tumor changes during neoadjuvant chemotherapy, Magn. Reson. Med. 73 (6) (2017) 2343–2356.
[20] D.E. Hurtado, N. Villarroel, C. Andrade, J. Retamal, G. Bugedo, A. Bruhn, Spatial patterns and frequency distributions of regional deformation in the healthy human lung, Biomech. Model. Mechanobiol. 16 (4) (2017) 1413–1423.
[21] X. Li, L.R. Arlinghaus, A.B. Chakravarthy, J. Farley, I.A. Mayer, V.G. Abramson, M. C. Kelley, I.M. Meszoely, J. Means-Powell, T.E. Yankeelov, Early DCE-MRI changes after longitudinal registration may predict breast cancer response to neoadjuvant chemotherapy, Biomed. Image Regist. 229–230 (2012).
[22] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, F. Prior, The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository, J. Digit. Imaging 26 (2013) 1045–1057.
[23] L.J. Esserman, C. Perou, M. Cheang, A. DeMichele, L. Carey, L.J. Van 't Veer, J. Gray, E. Petricoin, K. Conway, N. Hylton, D. Berry, Breast cancer molecular profiles and tumor response of neoadjuvant doxorubicin and paclitaxel: The I-SPY TRIAL (CALGB 150007/150012, ACRIN 6657), J. Clin. Oncol. 27 (18 Suppl) (2009) 515.
[24] A.D. Barker, C.C Sigman, G.J. Kelloff, N.M. Hylton, D.A. Berry, L.J. Esserman, I-SPY 2 : an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy, Nature 86 (2009) 97–100.
[25] C.A. Hudis, W.E. Barlow, J.P. Costantino, R.J. Gray, K.I. Pritchard, J.A. Chapman, J.A. Sparano, S. Hunsberger, R.A. Enos, R.D. Gelber, J.A. Zujewski, Proposal for standardized definitions for efficacy end points in adjuvant breast cancer trials:The STEEP System, J. Clin. Oncol. 25 (2007) 2127–2132, 252127-2132.
[26] J.G. Sled, A.P. Zijdenbos, A.C. Evans, A nonparametric method for automatic correction of intensity nonuniformity in MRI data, IEEE Trans. Med. Imaging 17 (1998) 87–97.
[27] M. Modat, G.R. Ridgway, Z.A. Taylor, M. Lehmann, J. Barnes, D.J. Hawkes, N. C. Fox, S. Ourselin, Fast free-form deformation using graphics processing units, Comput. Methods Programs Biomed. 98 (2010) 278–284.
[28] B.A. Ardekani, M. Braun, B.F. Hutton, I. Kanno, H Iida, A fully automatic multimodality image registration algorithm, J. Comput. Assisted Tomography 19 (1995) 615–623.
[29] Y. Diez, A. Gubern-Merida, L. Wang, S. Diekmann, J. Marti, B. Platel, J. Kramme, R. Marti, Comparison of methods for current-to-prior registration of breast DCE-MRI, Lecture Notes in Computer Sci. 8539 (2014) 689–695.
[30] D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, D.J. Hawkes, Nonrigid registration using free-form deformations : application to breast MR Images, IEEE Trans. Med. Imaging 18 (8) (1999) 712–721.
[31] J. Wu, Y. Ou, S.P. Weinstein, E.F. Conant, N. Yu, V. Hoshmand, B. Keller, A.B. Ashraf, M. Rosen, A. DeMichele, C. Davatzikos, Quantification of tumor changes during neoadjuvant chemotherapy with longitudinal breast DCE-MRI registration. *Medical Imaging 2015: Computer-Aided Diagnosis* 9414(2015), 502-508.
[32] B.B. Avants, C.L. Epstein, M. Grossman, J.C. Gee, Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain, Med Image Anal 12 (2008) 26–41.
[33] Y. Ou, A. Sotiras, N. Paragios, C. Davatzikos, DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting, Med. Image Anal. 15 (2011) 622–639.
[34] B.A. Ardekani, S. Guckemus, A. Bachman, M.J. Hoptman, M. Wojtaszek, J. Nierenberg, Quantitative comparison of algorithms for inter-subject registration of 3D volumetric brain MRI scans, J. Neurosci. Methods 142 (2005) 67–76.
[35] C. Davatzikos, S. Rathore, S. Bakas, S. Pati, M. Bergman, R. Kalarot, P. Sridharan, A. Gastounioti, N. Jahani, E. Cohen, H. Akbari, B. Tunc, J. Doshi, D. Parker, M. Hsieh, A. Sotiras, H. Li, Y. Ou, R.K. Doot, M. Bilello, Y. Fan, R.T. Shinohara, P. Yushkevich, R. Verma, D. Kontos, Cancer imaging phenomics toolkit : quantitative imaging analytics for precision diagnostics and predictive modeling of clinical outcome, J. Med. Imag. 5 (1) (2018), 011018.
[36] H. Uno, T. Cai, M.J. Pencina, R.B.D. Agostino, L.J Wei, On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data, Stat. Med. 30 (2011) 1105–1117.
[37] U. Sharma, K.K.A. Danishad, V. Seenu, N.R Jagannathan, Longitudinal study of the assessment by MRI and diffusion-weighted imaging of tumor response in patients with locally advanced breast cancer undergoing neoadjuvant chemotherapy, NMR Biomed 22 (2008) 104–113.
[38] C. Rousseau, A. Devillers, C. Sagan, L. Ferrer, B. Bridji, L. Campion, M. Ricaud, E. Bourbouloux, I. Doutriaux, M. Clouet, D. Berton-Rigaud, C. Bouriel, V. Delecroix, E. Garin, S. Rouquette, I. Resche, P. Kerbrat, J.F. Chatal, M. Campone, Monitoring of early response to neoadjuvant chemotherapy in stage II and III breast cancer by [18F]fluorodeoxyglucose positron emission tomography, J. Clin. Oncol. 24 (34) (2006) 5366–5372.
[39] M.S. Bae, S.U. Shin, H.S. Ryu, W. Han, S.A. Im, I.A. Park, D.Y. Noh, W.K. Moon, Pretreatment MR imaging features of triple-negative breast cancer: Association with response to neoadjuvant chemotherapy and recurrence-free survival, Radiology 281 (2016) 392–400.
[40] J. Lee, S.H. Kim, B.J Kang, Prognostic factors of disease recurrence in breast cancer using quantitative and qualitative magnetic resonance imaging (MRI) parameters, Sci. Rep. 10 (2020) 1–13, 7598.
[41] M.A. Mazurowski, L.J. Grimm, J. Zhang, P.K. Marcom, S.C. Yoon, C. Kim, S. V. Ghate, K.S. Johnson, Recurrence-free survival in breast cancer is associated with MRI tumor enhancement dynamics quantified using computer algorithms, Eur. J. Radiol. 84 (2015) 2117–2122.
[42] J.J. Doshi, G. Erus, Y. Ou, C. Davatzikos, Ensemble-based medical image labeling via sampling morphological appearance manifolds, MICCAI Chall. Work. Segmentation Algorithms, Theory Appl. (2013).
[43] S.E.A. Muenzing, B.V Ginneken, M.A. Viergever, J.P.W Pluim, DIRBoost-An algorithm for boosting deformable image registration: application to lung CT intra-subject registration, Med. Image Anal. 18 (2014) 449–459.