

Feasibility of contrast enhanced MRI derived textural features to predict overall survival in locally advanced breast cancer

Abstract

Background The prognosis for women with locally advanced breast cancer (LABC) is poor, and there is a need for better treatment stratification. Gray level co-occurrence matrix (GLCM) texture analysis of magnetic resonance (MR) images has been shown to predict pathological response, and could become useful in stratifying patients to more targeted treatments.

Purpose To evaluate the ability of GLCM textural features obtained prior to neoadjuvant chemotherapy to predict overall survival seven years post diagnosis of LABC patients.

Material and Method This retrospective study includes data from 55 LABC patients. GLCM textural features were extracted from segmented tumours in pre-treatment dynamic contrast enhanced 3T MR images. Prediction of overall survival by GLCM textural features was assessed and compared to predictions using traditional clinical variables.

Results Linear mixed-effect models showed significant differences in five GLCM features (f_1 , f_2 , f_5 , f_{10} , f_{11}) between survivors and non-survivors. Using discriminant analysis for prediction of survival, GLCM features from two minutes post-contrast images achieved a classification accuracy of 73% ($p < 0.001$), whereas traditional prognostic factors resulted in a classification accuracy of 67% ($p = 0.005$). Using a combination of both yielded the highest classification accuracy (78%, $p < 0.001$). Median values for features f_1 , f_2 , f_{10} and f_{11} provided significantly different survival curves in Kaplan-Meier analysis.

Conclusion This study shows a clear association between textural features from post-contrast images obtained prior to neoadjuvant chemotherapy, and overall survival seven years post diagnosis. Further studies in larger cohorts should be undertaken to investigate how this prognostic information can be used to benefit treatment stratification.

Introduction

Breast cancer is currently expected to constitute one third of all new cancer diagnoses in women (1). With emergence of better and more targeted treatments, mortality rates for female breast cancer has dropped. However, prognosis for patients diagnosed with locally advanced breast cancer (LABC) is still poor (2). Current treatment for LABC patients is multimodal, and includes chemotherapy, irradiation, endocrine therapy and surgical resection. Neoadjuvant chemotherapy (NAC) is often offered before surgery to down-stage tumours and improve operability, and to eradicate distant micro-metastases (3). NAC has been linked with improved long-term prognosis in patients who achieve pathological complete response (4, 5), but not all patients respond to treatment. Identifying prognostic factors, even before initiating NAC, would be beneficial for improved treatment stratification and overall survival in LABC patients.

Dynamic contrast-enhanced (DCE) magnetic resonance imaging (MRI) involves the acquisition of a baseline image and a series of images after the intravenous administration of a contrast agent. DCE-MRI allows three-dimensional visualisation of angiogenic properties in breast cancer, making it a powerful tool for detecting changes before morphological alterations (6). Pre-treatment MRI of the breast provides characteristics of the whole lesion and holds more comprehensive information on heterogeneity than biopsies, which represent only a small proportion of the tumour. DCE-MRI derived pharmacokinetic model parameters have been used in evaluating and predicting early response to NAC (7-12) as well as overall survival for patients treated with NAC (13). Another approach is Gray level co-occurrence matrix (GLCM) texture analysis, which quantifies spatial variations in gray-level intensity and provides information about intuitive qualities of the images, such as sharpness and homogeneity. Texture analysis of DCE-MRI has been used in breast cancer staging (14-19), for prediction of

invasive disease (20) and monitoring response to treatment (21-25), while a recent study (26) examined the potential to predict breast cancer survival using textural features from pre-treatment DCE-MRI along with shape features and traditional survival factors.

The main aim of this study was to assess the feasibility of GLCM features for prediction of 7-year overall survival in LABC patients, using the whole DCE time-series.

Materials and Methods

Patients and treatment

This retrospective study includes 56 Caucasian women with LABC treated with NAC, with available status for overall survival at least seven years after diagnosis. The study was approved by [anonymised – to be inserted later], and written informed consent was obtained from all patients. Patients were treated at [anonymised – to be inserted later] in the period 2007-2010 according to national guidelines (24). In brief, patients received four cycles of FEC, followed by four cycles of docetaxel, every 3 weeks. After the fourth cycle of FEC, the response to NAC was evaluated; if a decrease in the tumour longest diameter $\geq 80\%$ was achieved, the cycles of docetaxel were either cancelled or replaced by two cycles of FEC. After the last cycle of NAC, patients were appointed for mastectomy and axillary lymph node dissection with post-operative treatment according to national guidelines (24). Histopathological analysis of the resected breast mass and axillary nodes provided pathological response status. The hormone receptor status of the tumours were determined from the diagnostic biopsy obtained prior to onset of neoadjuvant chemotherapy, and classified as estrogen (ER) positive or progesterone (PgR) positive if $\geq 10\%$ of the cells stained positive. Human epidermal growth factor receptor 2 (HER-2) status was defined as positive by immunohistochemistry (score 3+) or by in situ

hybridization (gene amplification ratio>2) (24). We followed REporting recommendations for tumor MARKer (REMARK) prognostic criteria (27).

Imaging acquisition and processing

Patients were imaged prior to NAC on a 3T MR scanner (Siemens Tim Trio, Erlangen, Germany, software platform VB13A and VB15A) using a dedicated four-channel bilateral breast coil. The MRI acquisition and texture analysis are previously described [reference removed – to be inserted later]. In brief; T1-weighted DCE-MR images were acquired using a 3D radiofrequency-spoiled gradient echo sequence. One baseline image was acquired before administration of contrast (0.1mmol/kg) (Omniscan, GE Healthcare, Norway (47 patients), Magnevist, Bayer Healthcare Pharmaceuticals, USA (9 patients)). Seven post-contrast images were acquired with a temporal resolution of 1 minute. The imaging protocol changed after scanning 17 patients due to scanner software platform upgrade; larger flip angle (from 6° to 10°) and shorter repetition time (from 3.50 to 3.22 ms). All images were acquired with in-plane resolution of 1.1 mm² and slice thickness 1.1-1.5 mm. Images were corrected for motion artefacts (FSL package, Oxford FMRIB Centre, University of Oxford, Oxford, UK). The tumours were manually segmented on motion-corrected subtraction-images. Texture analysis was performed on the segmented tumours from all the slices containing tumour, from the baseline and post-contrast images, using two-dimensional GLCMs. Textural features from all slices were averaged, resulting in 16 GLCM features (f₁ through f₁₆) per patient per image. Processing of the images is described in detail in [reference removed – to be inserted later]. Representative MR images are shown in Fig. 1.

Statistical analysis

Linear mixed-effects models (LMM) were used to assess time-related changes in individual GLCM features related to overall survival, considering all post-contrast time points

simultaneously. LMM (28) were built by random-intercept models (R 3.1.1, function lme; ‘nlme’ package (29)) employing restricted maximum likelihood. Log-transformed GLCM features (individually for f_1 - f_{16}) from post-contrast images were used as response variables, while patient id was modelled as a random effect. Fixed effects included survival status, log-transformed pre-contrast GLCM features, MR protocol used (1 or 2), and time-point (minutes post contrast). Multiple testing correction was performed using the Benjamini-Hochberg method.

Multivariate analysis was performed by principal component analysis (PCA) and orthogonalized partial least squares discriminant analysis (OPLSDA) (PLS toolbox version 8.2.1, Eigenvector Research, Inc., Washington, USA) in MATLAB (MathWorks, Natick, MA, USA). PCA of autoscaled data was used to assess natural clusters of the data and to search for possible outliers. The benefit of applying OPLSDA is its capability to handle data sets with large numbers of variables compared to samples, and data sets where variables are correlated (collinearity), typical characteristics of variables derived from texture analysis. For discriminating survivors from non-survivors, separate OPLSDA models were built using GLCM features from the different time-points as input. Models were also built based on clinical prognostic factors alone, and on the combination of clinical prognostic factors and GLCM features from 2-minutes post-contrast images (the time-point which was found to best discriminate survivors/non-survivors by OPLSDA based on GLCM features alone). Clinical prognostic factors included were age at diagnosis, largest tumour diameter (measured by calliper prior to treatment), TNM stage, ER and PgR receptor status, HER2 status, type (invasive ductal carcinoma, invasive lobular carcinoma or other) and histological grade. For all models, the response variable was 7-year survival status. GLCM features and clinical prognostic factors were auto-scaled prior to modelling. Models were validated by 10-fold cross-validation, where the data set was split into 10 random subsets that were each kept out of

model-building and used for validation. The whole procedure was repeated with 20 iterations, with average classification results presented. The significance of the resulting model was evaluated using permutation testing with 1000 permutations; models with p-values (p_{perm}) <0.05 were considered significant. OPLSDA loading plots were coloured according to the variable's importance in the projection scores (VIP), reflecting the variable's influence on the classification. T-tests were used to assess if classification accuracies of OPLSDA models built using GLCM features only, clinical variables only, and a combination of both were significantly different, by comparing the distribution of classification accuracies from the 20 iterations.

The mean values of the individual GLCM features at the time-point that best discriminated between survivors and non-survivors (2 minutes post-contrast) were calculated. Differences in GLCM features between survivors/non-survivors at 2 minutes post-contrast were assessed by multiple linear regressions, correcting for the use of two acquisition sequences for DCE-MRI (MATLAB, MathWorks, Natick, MA, USA). GLCM features were log transformed prior to analysis to conform to normality.

Kaplan-Meier analysis (MATLAB, MathWorks, Natick, MA, USA) was used to depict differences in overall survival among patients with GLCM values above or below median derived from 2 minutes post-contrast images. GLCM features tested were significant in LMM analysis and/or important in OPLSDA analysis (VIP score >1.5); angular second moment (f_1), contrast (f_2), correlation (f_3), inverse difference moment (f_5), entropy (f_9), difference variance (f_{10}), difference entropy (f_{11}), and information measure of correlation 1 (f_{12}). Log-rank tests were performed to assess if the survival curves were significantly different.

To assess correlations between GLCM features, Pearson's correlation coefficients were calculated. A heat-map was established to visualize the correlations.

Results

Patient characteristics

One patient with multiple small and diffuse lesions was an outlier with high residual variation in PCA of GLCM features was excluded (Suppl. Fig. 1), resulting in a final cohort of 55 patients. Of these, 38 patients were survivors, while 17 patients were deceased within 7 years and classified as non-survivors. Table 1 presents clinical characteristics for the patient cohort. Median follow-up time was 106 months for survivors and 41 months for non-survivors.

Characteristics of textural features

Mean values and SD from 2 minutes post contrast are shown in Suppl. Table 1. A heat-map visualizing Pearson's correlation coefficients between the individual GLCM features at 2 minutes post-contrast is shown in Fig. 2. A strong and positive correlation was observed of GLCM textural features f_1 and f_5 , and of f_2 and f_9 , f_{10} and f_{11} . Simultaneously, f_1 and f_5 were negatively correlated with f_2 , f_9 , f_{10} and f_{11} .

Differences in GLCM values according to survival status

Table 2 shows the calculated p-values from LMM analysis assessing differences in GLCM values between survivors and non-survivors. LMM included GLCM textural features from all time-points simultaneously for assessment of significant differences using the whole DCE time-series. The GLCM features angular second moment (f_1), contrast (f_2), inverse difference moment (f_5), difference variance (f_{10}), and difference entropy (f_{11}) were significantly different between survivors and non-survivors before but not after multiple testing correction. Additionally, entropy (f_9) and cluster shade (f_{15}) approached significance. The change in the

MR protocol was found to significantly impact two of the textural features; variance (f_4) and cluster shade (f_{15}).

Textural features show added value to clinical prognostic factors in predicting overall survival.

Classification accuracies for survival based on multivariate analysis of GLCM textural features from different time-points post-contrast are shown in Fig. 3. Textural features provided classification accuracies ranging from 60-73%, with the highest accuracy (73.1%) obtained 2 minutes post contrast (sensitivity 65.3%, specificity 80.8%, $p_{perm}<0.001$). Scores and loadings from the OPLSDA model for predicting survival status from 2 minutes post-contrast images are shown in Fig. 4. Higher levels of features representing correlation (f_3) and inverse difference moment (f_5), and lower levels of features representing contrast (f_2), entropy (f_9), difference variance (f_{10}), difference entropy (f_{11}), and information measure of correlation 1 (f_{12}) from 2 minutes post-contrast in non-survivors were the most important differentiators. Clinical parameters provided a classification accuracy for prediction of survival of 67.2% (sensitivity 82.8%, specificity=51.7%, $p_{perm}=0.005$). The combination of clinical prognostic factors and GLCM features from 2 minutes post-contrast yielded a significantly higher ($p<0.001$) classification accuracy for survival compared to clinical parameters alone (accuracy 77.8%, sensitivity 79.0%, specificity 76.7%, $p_{perm}<0.001$).

Kaplan-Meier analysis shows correlation between GLCM textural features and survival.

Survival curves for f_1 , f_2 , f_9 , f_{10} and f_{11} were significantly different when dividing the patients in two groups using median values for GLCM textural features from 2 minutes post-contrast images as cut-off (p-values 0.015, 0.001, 0.041, 0.007, and 0.0015, respectively). Survival

curves for f_1 , f_2 , f_{10} , and f_{11} remained significant after multiple testing corrections (q-values < 0.05 , Fig. 5).

Discussion

In this paper, we have demonstrated the possibility to assess patient prognosis from pre-treatment DCE-MRI using GLCM textural features. Five out of 16 GLCM textural features were significantly different (prior to multiple testing) between survivors and non-survivors at 7 years of follow-up using LMM; angular second moment (f_1), contrast (f_2), inverse difference moment (f_5), difference variance (f_{10}) and difference entropy (f_{11}), while entropy (f_9) approached significance. Importantly, OPLSDA provided significant classification of survival status from 2 minutes post-contrast images, with the same GLCM textural features in addition to correlation (f_3) and information measure of correlation 1 (f_{12}) being the most important variables. GLCM textural features predicted overall survival by OPLSDA with accuracy comparable to clinical prognostic factors. Interestingly, while classification using clinical prognostic factors provided high sensitivity, GLCM textural features provided classification with high specificity. Combining GLCM features and clinical parameters provided the most accurate classification of survival, with both sensitivity and specificity approaching 80%. Grouping patients based on high and low values of textural features resulted in significant Kaplan-Meier survival curves for four of the same features, even when correcting for multiple testing.

Describing the underlying physiology that GLCM textural features represent is challenging. The highly positively correlated textural features f_1 and f_5 (Fig. 2) are in general associated with uniformity of gray-levels, with higher levels indicating a more homogeneous signal distribution within the tumour. In our cohort, non-survivors were associated with higher

uniformity. The textural features f_9 , f_{10} and f_{11} , also positively correlated, are associated with heterogeneity, and the non-survivors were associated with lower values of these features. In terms of physiology, this could point towards a fast, high and uniform uptake of contrast in the more aggressive tumours, resulting in more uniform signal distribution.

A few studies have assessed prognostic outcomes in breast cancer using pre-treatment DCE MRI data (13, 26, 30, 31). A strength of the current study is the 7-years follow-up, which allows the assessment of long-term outcomes. Our results are in accordance with the study of Kim et al. (31), who found that patients with breast cancers that appeared with less heterogeneity (lower entropy) on contrast enhanced T1-weighted subtraction images exhibited poorer outcomes. Further, Pickles et al. (26) evaluated the prognostic values of pre-treatment DCE-MRI-based parameters for breast cancer patients, including the same textural features as explored in our study. GLCM textural features representing heterogeneity of contrast enhancement; sum variance (f_7) and sum entropy (f_8), and symmetry of contrast enhancement; cluster shade (f_{15}) and cluster prominence (f_{16}), were consistently higher in patients with shorter survival (26). None of these features appeared as significantly different between the patient groups in our study, however, cluster shade (f_{15}) was found to be significantly affected by the change in the MR protocol, which could mask differences. Interestingly, one of the features with most significant association to survival in our cohort, f_{10} , (also related to heterogeneity) was previously shown to correlate with treatment response in the same dataset, with lower values reported for the non-responders [reference removed – to be inserted later]. An association between f_{10} and treatment response was also detected in an independent patient cohort (21), but then with higher values for the non-responders reported.

Several factors related to acquisition parameters and image processing can affect the robustness, reproducibility and reliability of textural features (32-34). This may explain why

studies report different textural features as important for similar classification purposes. Statistical analysis (LMM) showed differences by MR protocol in two of the features (Table 2), but these had limited importance in the classification of survival. This is supported by previous research showing that textural features are increasingly sensitive to acquisition parameters variation with increasing spatial resolution, but that the effect on pattern discrimination still is minimal provided sufficiently high spatial resolution (35). However, software platform upgrades are natural and necessary events in maintaining clinical scan protocols. Identifying robust features less affected by such changes could improve clinical translation of texture analysis. A different contrast agent was used for nine of the patients [reference removed – to be inserted later]. However, differences in image contrast due to different contrast agents have been found to be negligible (36), and we assumed this difference had no significant impact on the texture feature values. Prior to the extraction of textural features, we performed histogram equalization to minimize the effects of the differences in acquisition protocol. A previous study by Sikiö et al. (37) has shown some textural features to depend on the size of the segmented area used for the analysis. Larger tumour size is associated with a poorer prognosis in breast cancer, and the largest tumour diameter was significantly higher in the non-survivors of our cohort. However, none of the textural features correlated to tumour size in our study (results not shown).

Our study is based on a small patient cohort (N=55). A larger patient cohort would allow to construct more robust models, including validation using independent data. However, this study still demonstrates the feasibility of predicting patient prognosis in LABC using texture analysis of baseline DCE-MRI, before all clinical prognostic factors can be determined. Our results should be verified using larger cohorts and reproducibility of the textural features should be assessed, preferably using data from a multi-centre cohort to address the impact from differences in acquisition parameters.

A strength of our study is that the use of LMMs allowed information from the whole enhancement time-curve to be included in the analysis. However, the most significant associations for textural features and survival were still observed in the 2 minutes post-contrast images (Fig. 3), as also reported when associating textural features to treatment response (21, 24) demonstrating the clinical importance of using the textural features from the time-point where the contrast-uptake is at its peak. Also previous studies pinpoint the potential of the enhancement curve type to predict survival (30), where tumours with a high proportion of the aggressive curve type III voxels (38) were associated with a poorer prognosis. The fast and high contrast up-take in addition to the rapid wash out is the major features of this curve type.

In conclusion, textural features derived from contrast enhanced MRI prior to neoadjuvant chemotherapy in breast cancer patients are associated with 7-years survival outcome. Furthermore, the textural features show added value to the clinical prognostic factors in the prediction of long-term survival. This is in accordance with previous studies, where textural features have been associated with long-term outcome and treatment response, and shows a valuable potential to obtain more personalized treatment for breast cancer patients. However, our results need validation in a larger, independent cohort.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin* 2019;69:7-34.
2. Klein J, Tran W, Watkins E, et al. Locally advanced breast cancer treated with neoadjuvant chemotherapy and adjuvant radiotherapy: a retrospective cohort analysis. *BMC Cancer* 2019;19:306.
3. Gianni L, Eiermann W, Semiglazov V, et al. Neoadjuvant and adjuvant trastuzumab in patients with HER2-positive locally advanced breast cancer (NOAH): follow-up of a randomised controlled superiority trial with a parallel HER2-negative cohort. *Lancet Oncol* 2014;15:640-647.
4. Wang M, Hou L, Chen M, et al. Neoadjuvant Chemotherapy Creates Surgery Opportunities For Inoperable Locally Advanced Breast Cancer. *Sci Rep* 2017;7:44673.
5. Kong X, Moran MS, Zhang N, et al. Meta-analysis confirms achieving pathological complete response after neoadjuvant chemotherapy predicts favourable prognosis for breast cancer patients. *Eur J Cancer* 2011;47:2084-2090.
6. Turnbull LW. Dynamic contrast-enhanced MRI in the diagnosis and management of breast cancer. *NMR Biomed* 2009;22:28-39.

7. Blinded

8. Cheung YC, Chen SC, Su MY, et al. Monitoring the size and response of locally advanced breast cancers to neoadjuvant chemotherapy (weekly paclitaxel and epirubicin) with serial enhanced MRI. *Breast Cancer Res Treat* 2003;78:51-58.

9. Hylton NM, Blume JD, Bernreuter WK, et al. Locally advanced breast cancer: MR imaging for prediction of response to neoadjuvant chemotherapy--results from ACRIN 6657/I-SPY TRIAL. *Radiology* 2012;263:663-672.

10. Li X, Arlinghaus LR, Ayers GD, et al. DCE-MRI analysis methods for predicting the response of breast cancer to neoadjuvant chemotherapy: pilot study findings. *Magn Reson Med* 2014;71:1592-1602.

11. Drisis S, Metens T, Ignatiadis M, et al. Quantitative DCE-MRI for prediction of pathological complete response following neoadjuvant treatment for locally advanced breast cancer: the impact of breast cancer subtypes on the diagnostic accuracy. *Eur Radiol* 2016;26:1474-1484.

12. Drew PJ, Kerin MJ, Mahapatra T, et al. Evaluation of response to neoadjuvant chemoradiotherapy for locally advanced breast cancer with dynamic contrast-enhanced MRI of the breast. *Eur J Surg Oncol* 2001;27:617-620.

13. Blinded

14. Cai H, Peng Y, Ou C, et al. Diagnosis of breast masses from dynamic contrast-enhanced and diffusion-weighted MR: a machine learning approach. *PLoS One* 2014;9:e87387.
15. Chen W, Giger ML, Li H, et al. Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images. *Magn Reson Med* 2007;58:562-571.
16. Karahaliou A, Vassiou K, Arikidis NS, et al. Assessing heterogeneity of lesion enhancement kinetics in dynamic contrast-enhanced MRI for breast cancer diagnosis. *Br J Radiol* 2010;83:296-309.
17. Milenkovic J, Hertl K, Kosir A, et al. Characterization of spatiotemporal changes for the classification of dynamic contrast-enhanced magnetic-resonance breast lesions. *Artif Intell Med* 2013;58:101-114.
18. Nagarajan MB, Huber MB, Schlossbauer T, et al. Classifying Small Lesions on Breast MRI through Dynamic Enhancement Pattern Characterization. *Mach Learn Med Imaging* 2011;7009:352-359.
19. Nagarajan MB, Huber MB, Schlossbauer T, et al. Classification of small lesions in dynamic breast MRI: eliminating the need for precise lesion segmentation through spatio-temporal analysis of contrast enhancement. *Mach Vis Appl* 2013;24:1371-1381.

20. Harowicz MR, Saha A, Grimm LJ, et al. Can Algorithmically Assessed MRI Features Predict Which Patients With a Preoperative Diagnosis of Ductal Carcinoma In Situ Are Upstaged to Invasive Breast Cancer? *J Magn Reson Imaging* 2017;46:1332-1340.
21. Ahmed A, Gibbs P, Pickles M, et al. Texture analysis in assessment and prediction of chemotherapy response in breast cancer. *J Magn Reson Imaging* 2013;38:89-101.
22. Golden DI, Lipson JA, Telli ML, et al. Dynamic contrast-enhanced MRI-based biomarkers of therapeutic response in triple-negative breast cancer. *J Am Med Assoc* 2013;20:1059-1066.
23. Michoux N, Van den Broeck S, Lacoste L, et al. Texture analysis on MR images helps predicting non-response to NAC in breast cancer. *BMC Cancer* 2015;15:574-587.
24. Blinded
25. Wu J, Gong G, Cui Y, et al. Intratumor partitioning and texture analysis of dynamic contrast-enhanced (DCE)-MRI identifies relevant tumor subregions to predict pathological response of breast cancer to neoadjuvant chemotherapy. *J Magn Reson Imaging* 2016;44:1107-1115.
26. Blinded
27. McShane LM, Altman DG, Sauerbrei W, et al. REporting recommendations for tumor MARKer prognostic studies (REMARK). *Breast Cancer Res Treat* 2006;100:229-235.

28. Pinheiro J, Bates D. Linear Mixed-Effects Model: Basic Concepts and Examples. In: Pinheiro J, Bates D, editors. *Mixed-Effects Models in S and S-PLUS*. New York: Springer; 2000.
29. Pinheiro J, Bates D, DebRoy S, et al. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-117. 2014.
30. Tuncbilek N, Tokatli F, Altaner S, et al. Prognostic value DCE-MRI parameters in predicting factor disease free survival and overall survival for breast cancer patients. *Eur J Radiol* 2012;8:863-867.
31. Kim JH, Ko ES, Lim Y, et al. Breast Cancer Heterogeneity: MR Imaging Texture Analysis and Survival Outcomes. *Radiology* 2017;282:665-675.
32. Scalco E, Rizzo G. Texture analysis of medical images for radiotherapy applications. *Br J Radiology* 2017;90: 20160642
33. Collewet G, Strzelecki M, Mariette F. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magn Reson Imaging* 2004;22:81-91.
34. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017;14:749-762.

35. Mayerhoefer ME, Szomolanyi P, Jirak D, et al. Effects of MRI acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination: an application-oriented study. *Med Phys* 2009;36:1236-1243.
36. Kanal E, Maravilla K, Rowley HA. Gadolinium Contrast Agents for CNS Imaging: Current Concepts and Clinical Evidence. *Am J Neuroradiol* 2014;35:2215-2226.
37. Sikio M, Holli KK, Harrison LCV, et al. Parkinson's Disease: Interhemispheric Textural Differences in MR Images. *Acad Radiol* 2011;18:1217-1224.
38. Kuhl CK, Mielcareck P, Klaschik S, et al. Dynamic breast MR imaging: Are signal intensity time course data useful for differential diagnosis of enhancing lesions? *Radiology* 1999;211:101-110.

Table 1 Description of the patient cohort. Clinical characteristics before administration of neoadjuvant chemotherapy for survivors (N=38) and non-survivors (N=17) of locally advanced breast cancer.

	Survivors (N = 38)	Non-survivors (N =17)	p
Mean age (range) at diagnosis (years)	56 (35-82)	52 (35-73)	0.239 ^d
Largest tumour dimension (millimeters)	62 (30-100)	70 (30-100)	0.010^d
Clinical stage^a			< 0.001^c
IIB (T3 N0 M0)	17 (45%)	5 (30%)	
IIIA (T2-3 N1-2 M0)	12 (32%)	2 (11%)	
IIIB (T4a-d N0-2 M0)	9 (23%)	3 (18%)	
IIIC (T1-2 N3 M0)	0	1 (6%)	
IV (T3-4c N1-2 M1)	0	6 (35%)	
T			0.069 ^c
2	1 (3%)	1 (6%)	
3	28 (74%)	10 (59%)	
4	9 (23%)	6 (35%)	
N			0.008^c
0	21 (55%)	6 (35%)	
1	13 (34%)	6 (35%)	
2	4 (11%)	4 (24%)	
3	0	1 (6%)	
M			< 0.001^c
0	38 (100%)	11 (65%)	
1	0	6 (35%)	
Receptor status^b			
ER+	25 (66%)	13 (76%)	0.749 ^c
PgR+	17 (45%)	7 (41%)	0.777 ^c
HER2+	14 (37%)	3 (18%)	0.209 ^c

Unknown	1 (3%)	0	
Histopathological type of breast cancer ^c			0.784 ^c
IDC	35 (92%)	15 (88%)	
ILC	2 (5%)	1 (6%)	
Mucinous carcinoma	1 (3%)	1 (6%)	
Grade of breast cancer			0.127 ^c
Grade 1	3 (8%)	1 (6%)	
Grade 2	19 (50%)	4 (23%)	
Grade 3	16 (42%)	12 (71%)	
Response to treatment			0.543 ^c
Stable disease	13 (34%)	5 (30%)	
Partial response	12 (32%)	8 (47%)	
Complete response	13 (34%)	4 (23%)	
Presence of lymph node metastasis	17 (45%)	11 (65%)	0.245 ^c

^a

T2: Tumour is more than 2 cm but no more than 5 cm across.

T3: Tumour is bigger than 5 cm across.

T4: Tumour has either spread to the skin, the chest wall or both, or inflammatory carcinoma is present.

N0: No nearby lymph node metastasis.

N1-2: Metastasis to movable/fixed ipsilateral axillary nodes

N3: Metastasis to ipsilateral internal mammary lymph nodes

M0: There is no distant metastasis.

M1: Distant metastasis is present.

^b ER+: Estrogen receptor negative

PgR+: Progesterone receptor positive HER2+: Herceptin receptor positive ^c IDC = invasive ductal carcinoma,

ILC = invasive lobular carcinoma

^d p-value was calculated using independent t-test for two variables.

^e p-value was calculated using Fisher's exact test.

Table 2 GLCM textural features with significantly different levels between survivors and non-survivors. Linear mixed-effects model p-values for the relationship of grey level co-occurrence matrix features from all time-points to survival in locally advanced breast cancer patients are shown. p-values below 0.05 are presented in bold. No features were significant after Benjamini-Hochberg correction for multiple testing.

Feature	p-value
Angular second moment (f_1)	0.008
Contrast (f_2)	0.034
Correlation (f_3)	0.308
Variance (f_4)	0.787 [†]
Inverse difference moment (f_5)	0.008
Sum average (f_6)	0.122
Sum variance (f_7)	0.470
Sum entropy (f_8)	0.366
Entropy (f_9)	0.050
Difference variance (f_{10})	0.024
Difference entropy (f_{11})	0.018
Information measure of correlation 1 (f_{12})	0.347
Information measure of correlation 2 (f_{13})	0.777
Maximal correlation coefficient (f_{14})	0.358
Cluster shade (f_{15})	0.087 [†]
Cluster prominence (f_{16})	0.477

[†] Feature affected by change in MR protocol

Figure legends

Fig. 1 MR image examples. 2 minutes post-contrast dynamic contrast-enhanced MR images of (A) a survivor and (B) a non-survivor of locally advanced breast cancer. The tumor region is outlined in red. While the segmentation was performed on the subtracted images, the figure shows the tumour volume in non-subtracted 2 minutes post-contrast images for better visualization.

Fig. 2 Heat-map of correlations between textural features. The heat-map based on Pearson correlations between the textural features illustrates correlations, ranging from strong negative (-1) to strong positive (+1), between grey level co-occurrence matrix features f1 to f16 obtained from 2 minutes post-contrast images.

Fig. 3 Classification accuracy for the prediction of survival using orthogonalized partial least squares discriminant analysis at each timepoint. Classification accuracy by partial least squares discriminant analysis using all grey level co-occurrence matrix (GLCM) features obtained from post-contrast images of locally advanced breast cancer patients (N=55) at each time-point (0 to 7 minutes post-contrast). The circle marks the maximum classification accuracy, which is at 2 minutes post-contrast.

Fig. 4 Scores and loadings from orthogonalized partial least squares discriminant analysis (OPLSDA) for overall survival. Scores (A) and loadings (B) on latent variable (LV) 1 from OPLSDA predicting survival using as input all grey level co-occurrence matrix (GLCM) features obtained from 2 minutes post-contrast images of locally advanced breast cancer patients (N=55). In the loadings, variables are coloured according to their importance for projection (VIP) score, reflecting the variable's importance for the classification. For this

cohort, OPLSDA classified survivors and non-survivors with an accuracy of 73.1% ($p_{\text{perm}} < 0.001$).

Figure 5 Kaplan Meier survival curves for patients with high and low values of selected grey level co-occurrence matrix (GLCM) features. The recurrence-free proportions are plotted against overall survival time for GLCM features dichotomized to above or below median values. Only survival curves for features that f1, f2, f10, and f11 remained significant after multiple testing corrections are shown (f1, f2, f10, and f11)

Supporting information captions

Supplementary Figure 1 Scores from principal component analysis indicate the outlier.

Scores from principal component (PC) analysis of grey level co-occurrence matrix features obtained from 2 min post-contrast images. The outlier is shown in red. This patient was characterized with multiple small and diffuse lesions, and was excluded from the further analysis.

Supplementary Table 1 Mean and standard deviation of 16 GLCM feature values from 2 min post-contrast images comparing survivors and non-survivors. p-values were calculated using multiple linear regression. Significant results at 0.05 significance are shown in bold. No p-values were significant after Benjamini-Hochberg correction.

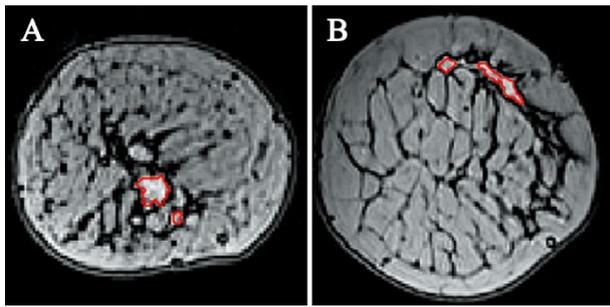


Figure 1

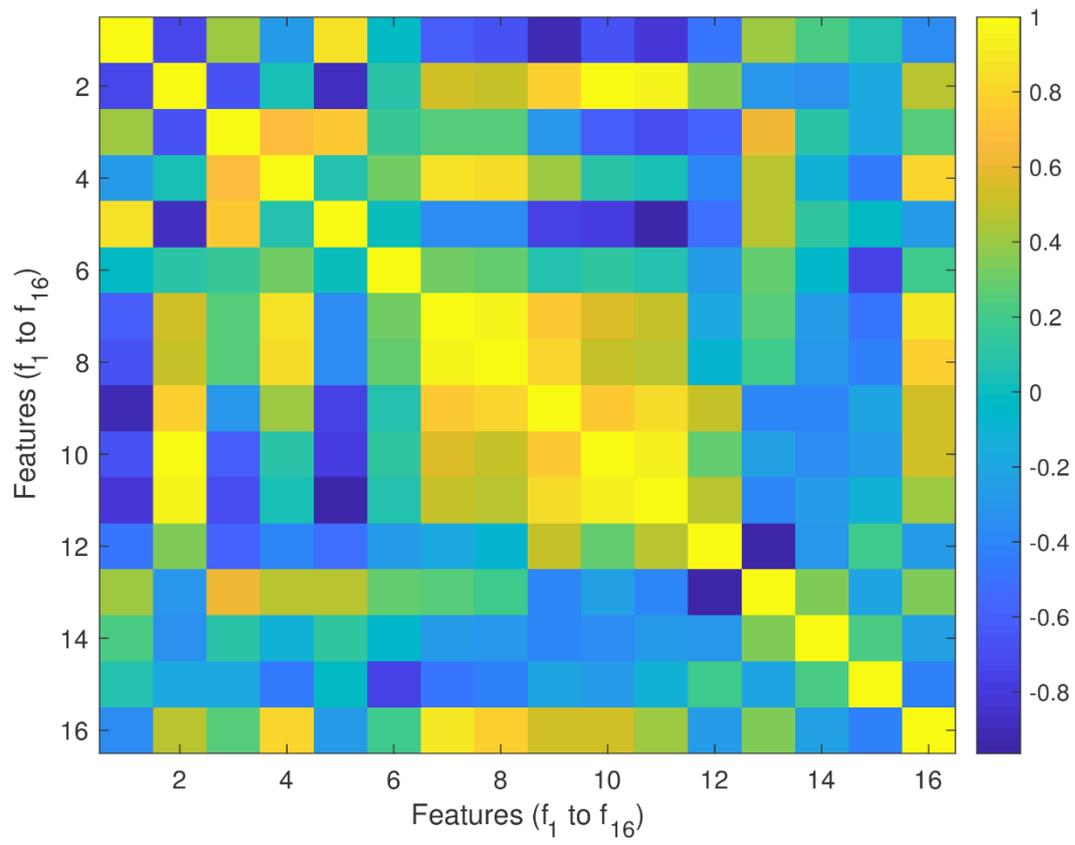


Figure 2

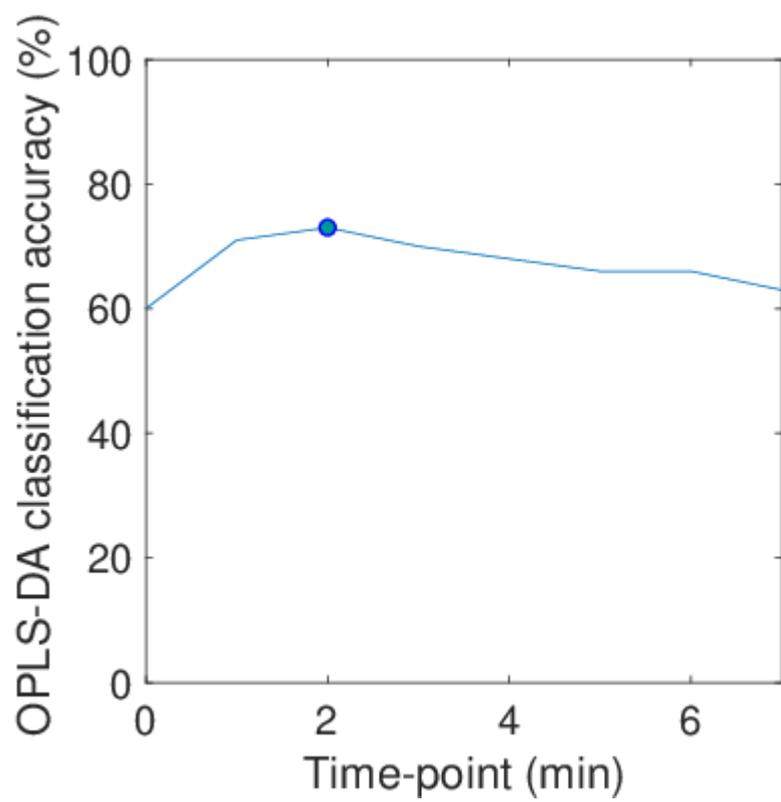


Figure 3

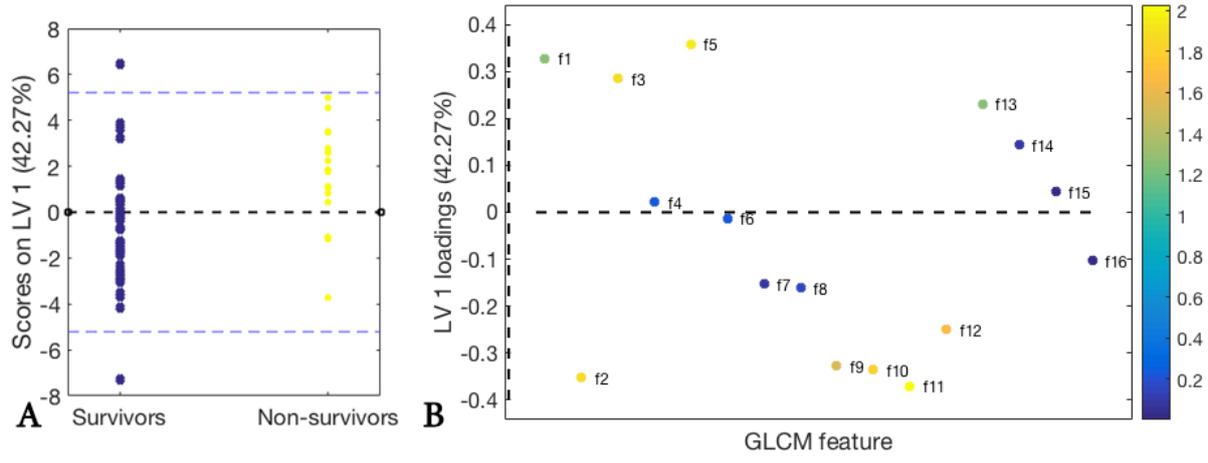


Figure 4

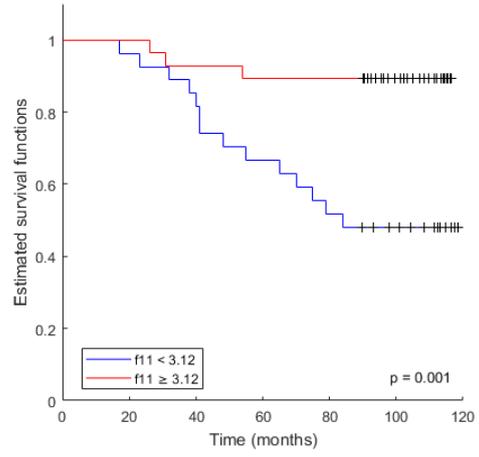
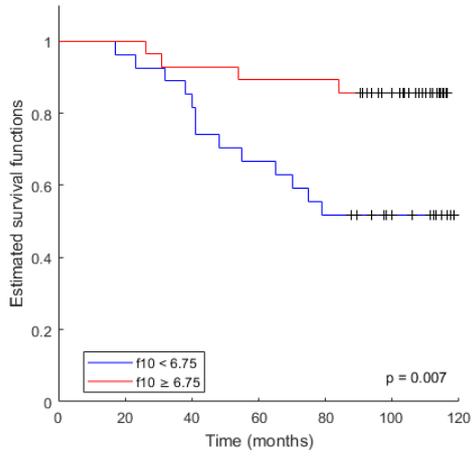
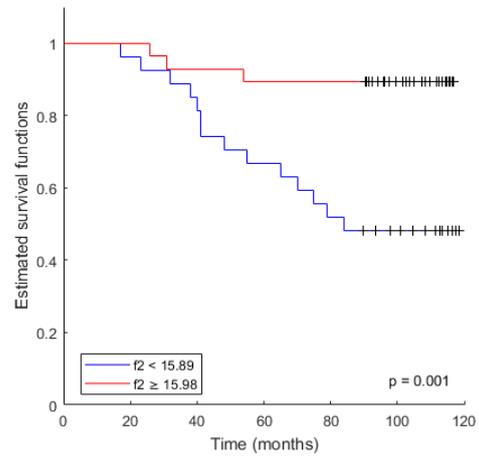
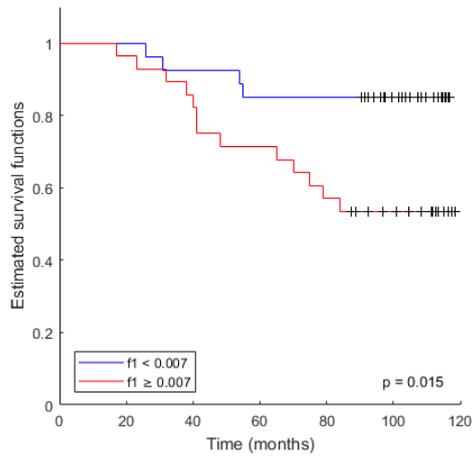
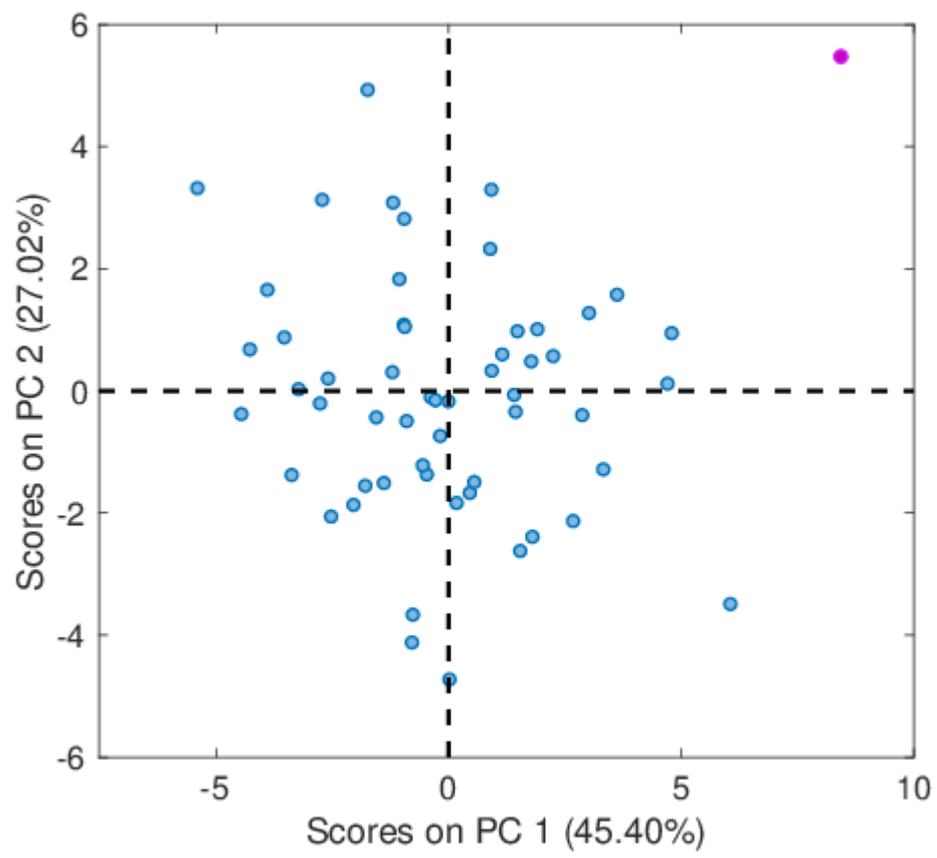


Figure 5



Supplementary Figure 1

Supplementary Table 1 Mean and standard deviation of 16 GLCM feature values from 2 min post-contrast images comparing survivors and non-survivors. p-values were calculated using multiple linear regression. Significant results at 0.05 significance are shown in bold. No p-values were significant after Benjamini-Hochberg correction for multiple testing.

Feature		Overall survival (N=55)	
		2-minutes post-contrast	p
Angular second moment (f_1)	Survivors	0.007 ± 0.003	0.022
	Non-survivors	0.009 ± 0.002	
Contrast (f_2)	Survivors	18.37 ± 6.57	0.006
	Non-survivors	13.65 ± 4.33	
Correlation (f_3)	Survivors	0.48 ± 0.11	0.010
	Non-survivors	0.56 ± 0.10	
Variance (f_4)	Survivors	8.28 ± 2.75	0.300
	Non-survivors	8.99 ± 2.86	
Inverse difference moment (f_5)	Survivors	0.29 ± 0.05	0.005
	Non-survivors	0.34 ± 0.05	
Sum average (f_6)	Survivors	37.63 ± 3.59	0.293
	Non-survivors	36.52 ± 5.43	
Sum variance (f_7)	Survivors	51.58 ± 13.66	0.703
	Non-survivors	49.66 ± 12.05	
Sum entropy (f_8)	Survivors	4.82 ± 0.18	0.466
	Non-survivors	4.78 ± 0.19	
Entropy (f_9)	Survivors	7.71 ± 0.37	0.021
	Non-survivors	7.46 ± 0.32	
Difference variance (f_{10})	Survivors	7.87 ± 2.48	0.009
	Non-survivors	6.12 ± 1.67	
Difference entropy (f_{11})	Survivors	3.18 ± 0.25	0.007
	Non-survivors	2.98 ± 0.20	
Information measure of correlation 1 (f_{12})	Survivors	-0.09 ± 0.03	0.015
	Non-survivors	-0.12 ± 0.04	
Information measure of correlation 2 (f_{13})	Survivors	0.70 ± 0.09	0.044
	Non-survivors	0.76 ± 0.09	
Maximal correlation coefficient (f_{14})	Survivors	0.96 ± 0.13	0.474
	Non-survivors	0.99 ± 0.09	
Cluster shade (f_{15})	Survivors	-56.8 ± 140.3	0.575
	Non-survivors	-35.0 ± 186.4	
Cluster prominence (f_{16})	Survivors	7972.5 ± 3411.9	0.826
	Non-survivors	8011.5 ± 3052.0	

