

Title: **Support vector machine for breast cancer classification using diffusion-weighted MRI histogram features: preliminary study**

Author Names and Degrees:

Igor Vidić, MSc¹, Liv Egnell, MSc^{1,2}, Neil P. Jerome, PhD^{2,3}, Jose R. Teruel, PhD^{4,5},
Torill E. Sjøbakk, PhD³, Agnes Østlie, MD², Hans E. Fjøsne, MD^{6,7}, Tone F. Bathen,
PhD³, Pål Erik Goa, PhD^{1,2}

Affiliations

¹Department of Physics, NTNU - Norwegian University of Science and Technology, Trondheim, Norway.

²Clinic of Radiology and Nuclear Medicine, St. Olavs University Hospital, Trondheim, Norway.

³Department of Circulation and Medical Imaging, NTNU - Norwegian University of Science and Technology, Trondheim, Norway.

⁴Department of Radiology, University of California San Diego, La Jolla, CA, United States

⁵Department of Radiation Oncology, NYU Langone Medical Center, New York, NY, USA

⁶Department of Cancer Research and Molecular Medicine, NTNU - Norwegian University of Science and Technology, Trondheim, Norway.

⁷Department of Surgery, St. Olavs University Hospital, Trondheim, Norway.

Corresponding Authors Info:

Pål Erik Goa, Department of Physics, Norwegian University of Science and Technology, Realfagbygget 7491 Trondheim, Norway. Tel: +47 73593634. Email: pal.e.goa@ntnu.no

Igor Vidić, Department of Physics, Norwegian University of Science and Technology, Realfagbygget 7491 Trondheim, Norway. Tel: +47 48331338. Email: igor.vidic@ntnu.no

Grant Support: NTNU - Norwegian University of Science and Technology, Trondheim, Physics department, no external grant.

Running Title:

Machine learning in DWI of breast cancer

ABSTRACT

BACKGROUND: Diffusion-weighted MRI (DWI) is currently one of the fastest developing MRI-based techniques in oncology. Histogram properties from model fitting of DWI are useful features for differentiation of lesions, and classification can potentially be improved by machine learning.

PURPOSE: To evaluate classification of malignant and benign tumors and breast cancer subtypes using Support Vector Machine (SVM).

STUDY TYPE: Prospective.

SUBJECTS: Fifty-one patients with benign (n=23) and malignant (n=28) breast tumors (26 ER+ whereof 6 were HER2+).

FIELD STRENGTH/SEQUENCE: Patients were imaged with DW-MRI (3T) using twice refocused SE-EPI with TE/TR=9000/86ms, 90x90 matrix size, 2x2mm in-plane resolution, 2.5mm slice thickness, and 13 b-values.

ASSESSMENT: Apparent Diffusion Coefficient (ADC), Relative Enhanced Diffusivity (RED) and the Intravoxel Incoherent Motion (IVIM) parameters diffusivity (D), pseudo-diffusivity (D*) and perfusion fraction (f) were calculated. The histogram properties (median, mean, standard deviation, skewness, kurtosis) were used as features in SVM (10-fold cross-validation) for differentiation of lesions and subtyping.

STATISTICAL TESTS: Accuracies of the SVM classifications were calculated to find the combination of features with highest prediction accuracy. Mann-Whitney tests were performed for univariate comparisons.

RESULTS: For benign versus malignant tumors, univariate analysis found 11 histogram properties to be significant differentiators. Using SVM, highest accuracy (0.96) was achieved from a single feature (mean of RED), or from 3 feature-combinations of IVIM or ADC. Combining features from all models gave perfect classification. No single feature predicted HER2 status of ER+ tumors (univariate or SVM), although high accuracy (0.90) was achieved with SVM combining several features. Importantly, these features had to include higher order statistics (kurtosis and skewness), indicating the importance to account for heterogeneity.

DATA CONCLUSION: Our findings suggest that SVM, using features from a combination of diffusion models, improves prediction accuracy for differentiation of benign versus malignant breast tumors, and may further assist in subtyping of breast cancer.

Keywords: Breast MR; Diffusion weighted MRI; Intravoxel incoherent motion; Support vector machine; Tumor heterogeneity; Prognostic factors

INTRODUCTION

Diffusion-Weighted Magnetic Resonance Imaging (DW-MRI) [1] is sensitive to the random motion of water molecules. In the case of Gaussian, or free diffusion, the DW-MRI signal can be appropriately modelled as a mono-exponential decay that is a function of the degree of applied diffusion weighting (b-value) and the diffusion coefficient. In cases of non-Gaussian diffusion, the mono-exponential model is commonly applied to data arising from application of two or more low and medium b-values, providing the measure of an apparent diffusion coefficient (ADC) that is influenced by number and choice of b-values [2].

It has been shown that ADC is able to detect cancerous tissue in breast, with reduced ADC being associated with malignant tumors [3-5]. The micro-structural origin of this effect is commonly attributed to increased cellular density and decreased extracellular matrix in cancer tissue compared to healthy fibroglandular tissue and benign lesions [3, 6].

At low b-values ($<50 \text{ s/mm}^2$), DW-MRI signal attenuation may also reflect blood flow in randomly-oriented capillaries, with this component having a pseudo-diffusion coefficient similar to that observed for true diffusion effects, but approximately one order of magnitude higher [7]. The bi-exponential intra-voxel incoherent motion (IVIM) [7] model is formulated to capture the combined effect of microcirculation (perfusion, f and D^*) in capillaries alongside conventional true diffusivity (D), and has been successfully applied to breast cancer [8, 9], providing accurate identification of malignant lesions.

Relative Enhanced Diffusivity (RED) [10] was recently introduced as a new approach in classification of breast lesions using DW-MRI data. RED is sensitive both to diffusion

and microcirculation, and has been shown to correlate with early enhancement in Dynamic Contrast Enhanced (DCE) MRI, an indicator of high perfusion within the lesion [10]. Compared with IVIM, RED is a simpler approach, requiring fewer b-values and thereby less data and faster to acquisition. This is at the cost of not providing quantitative biophysical parameters like the pseudo-diffusion (D^*) coefficient or pseudo-diffusion fraction (f), i.e., it does not strictly separate microvasculature induced diffusion from true diffusion.

While it is common to report the median/mean value of derived model parameters [11, 12] as measured across a region of interest (ROI), more detailed information can potentially be obtained from analysis of the parameter histograms [13]. These additional metrics provide greater insight into heterogeneity of tumors, which is important for the optimal planning of treatment [14]. MRI histogram features have already been associated with known prognostic factors for treatment outcome, including estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2) status [15].

Breast cancer is one of the few cancer types in which molecular classification (ER, PR, HER2) has successfully been used for the design of individualized therapies, leading to significant improvements in survival [16, 17]. The current method for determining the molecular subtype, however, currently requires invasive biopsies for histologic evaluation. Access to non-invasive prognostic and diagnostic factors derived from DW-MRI would further improve the treatment of breast cancer.

In addition to simple summary statistics, combining features from different MRI models can lead to improved diagnostic classification compared to single metrics or models [18,

19]. For such purposes it is common to use a machine learning approach, such as Support Vector Machine (SVM) [20]. SVM has been used for prediction in breast cancer classification [18, 19, 21], and showed the highest prediction accuracy among other machine learning methods.

The study aimed to establish whether the performance of DW-MRI can be improved using SVM with a combination of diffusion-derived parameters, specifically to differentiate malignant from benign tumors, and further to predict the HER2 status in ER positive breast cancer.

MATERIALS AND METHODS

Patient Cohort

The study was approved by the Regional Committee for Medical and Health Research Ethics (REK Central Norway, 2011/568). All patients gave written informed consent prior to enrolment. The recruitment of patients for this study started in October 2013 and ended in August 2016.

Following MR examination, patients with malignant tumors underwent surgery and histopathologic analysis was performed on the resected mass. ER status was classified as positive if $\geq 1\%$ of the cells were stained positive [22]. HER2 status was assessed by fluorescence in situ hybridization and classified as positive if the HER2 gene to chromosome ratio was ≥ 2.0 [23]. Categorization of benign tumors was done by histopathologic analysis on core needle biopsies or on resected tissue if the tumor was surgically removed. For benign lesions where biopsy was not requested by the radiologist, diagnosis was based on the patient history, which included either

radiographic mammography, ultrasonography, or a previous clinical MR examination with at least 6 months' follow-up at the time of recruitment. MR was performed on 61 patients; 10 datasets were excluded from analysis (7 non-successful motion correction and 3 had Nyquist ghosting artefacts), giving 51 cases in total. Where multiple lesions were present in the same breast, the largest was selected for analysis.

Of the 51 patients, 23 tumors were classified as benign and 28 as malignant. Most of the patients with malignant tumors were ER+ (N=26) and these were used for the classification of HER2 status, whereof 6 were HER2+ and 20 HER2- cases. Clinical data is reported in Table 1. A subset of thirty four out of the 51 patients analyzed in this study were previously reported in a study by Teruel et al. where the RED parameter was presented for the first time [10].

MRI Protocols

Patients were imaged with a 3T scanner (Skyra, Siemens Healthcare, Erlangen, Germany) equipped with a 16-channel breast coil (16-channel AI Breast Coil, Siemens Healthcare, Erlangen, Germany).

Fat-suppressed (n=17 FatSat and n=34 SPAIR) unilateral sagittal DWI was acquired using a twice-refocused spin-echo echo-planar imaging sequence with: repetition time (TR) 9000ms, echo time (TE) 86ms 90x90 matrix, 2x2mm in-plane resolution, slice thickness 2.5mm, 60 slices, generalized auto-calibrating partially parallel acquisition (GRAPPA) factor 2 and 13 b-values: 0, 10, 20, 30, 40, 50, 70, 90, 120, 150, 200, 400, 700s/mm² in six (n=18 (7 benign, 8 ER+HER2-, 1 ER+HER2+ and 2 ER-PR-HER2+), scan time 11 minutes) or three (n=33 (16 benign, 12 ER+HER2- and 5 ER+HER2+),

scan time 6 minutes) directions. The protocol included one additional geometry-matched, non-diffusion-weighted ($b=0\text{s/mm}^2$) series with reversed phase-encoding direction for implementation of distortion correction arising from susceptibility boundaries [24]. Twice-refocused diffusion encoding scheme was chosen to minimize eddy current effects [25]. The patients also underwent dynamic contrast enhanced (DCE) MRI. DCE scans consisted of 3D, T1 weighted, non-fat suppressed, gradient echo sequence (TR/TE 5.82/2.18, flip angle 15%, 256x256 matrix, in-plane resolution 0.7x0.7mm, slice thickness 2.5mm) acquisitions, collected pre-contrast, and at 7 consecutive time points (with temporal resolution of 1min) after administration of contrast agent. The DCE MR images were used for guidance of region of interest (ROI) selection in the DWI images.

For T2 weighted images, non-fat suppressed 2D turbo spin echo was performed, with TR/TE 5500/118ms, 256x256 matrix in-plane resolution 0.7x0.7mm, and slice thickness 2.5mm.

Data Analysis

Preprocessing

The processing workflow is presented in Figure 1; Before statistical analyses and machine learning, images were corrected for geometric distortion using the phase-reversed $b=0\text{s/mm}^2$ acquisition, using the method described by Holland et al [26] and proved for breast applications by Teruel et al [24]. In the case of displayed obvious patient motion, three-dimensional rigid co-registration using a normalized cross-

correlation metric was performed [27] coregistering all raw images to the corresponding b0 image. Following distortion and/or motion corrections, trace images were calculated and lesions were segmented in 3 dimensions on the largest b-value DW-MRI (700s/mm²) with reference to DCE images. ROIs were drawn by a basic scientist (I.V. with 2 years of experience in breast imaging) advised by a breast radiologist (A.Ø., with 20 years of experience).

Diffusion Models

ADC was calculated by fitting (Trust-Region method) the mono-exponential decay for the signal at b-values of 200, 400, and 700s/mm²:

$$\frac{S(b)}{S(b = 200)} = e^{-(b-200) \cdot ADC}$$

The RED parameter presents relative increase in ADC for lower b-values compared to the medium b-value range. It was calculated using the formulation by Teruel et al[10]:

$$RED = \frac{ADC_{b0,b1} - ADC_{b1,b2}}{ADC_{b1,b2}}$$

Where b0, b1, and b2 are 0, 200 and 700 s/mm², respectively.

The IVIM parameters true diffusivity (D), pseudodiffusion fraction (f), and pseudodiffusivity (D^*) were calculated from the bi-exponential IVIM model [7]:

$$\frac{S_b}{S_0} = (1 - f) \cdot e^{-b \cdot D} + f \cdot e^{-b(D+D^*)}$$

Since it can be assumed that the contribution to the signal coming from blood flow is negligible for $b > 200$ s/mm², fitting was performed using the segmented approach [8]; D was calculated from the monoexponential decay for the b-values higher than 200s/mm²

(400 and 700s/mm²). The zero intercept (S_{int}) of that monoexponential decay was used for estimation of the pseudodiffusion fraction parameter f :

$$f = \frac{S(b = 0) - S_{int}}{S(b = 0)}$$

Finally, D^* was calculated by constrained (0-100 μ m²/ms) Trust-region fitting to the bi-exponential IVIM equation fixing these values for D and f .

The distortion and motion corrections were performed using the preprocessing algorithm provided in the Computational Morphometry Toolkit (CMTK, SRI International, Menlo Park, CA). Image analysis and fitting were performed using in-house developed scripts in Matlab (2014a Mathworks, Natick).

Statistical Analysis

For each 3D ROI, voxel-wise fitting for ADC, RED and IVIM was performed to provide parameter maps of the tumors. For each of these parameters (ADC, RED, and D , f and D^* from IVIM), the mean, standard deviation, median, skewness, and kurtosis were calculated, giving 5 features for each of the 5 parameters and thus a total of 25 features for the machine learning analysis (Figure 1).

To utilize these features, the support vector machine architecture [20] was employed as shown schematically in Figure 2. The radial basis function was chosen as kernel, with kernel size (γ) and regularization parameter (C). Kernel size (γ) determines the range of influence of samples selected by the model as support vectors, while the regularization parameter (C) determines the number of selected support vectors, defining the complexity (smoothness) of the decision surface. A grid search over all possible

combinations of γ and C , within a preselected range of values ($\gamma, C \in [0.01, 0.1, 0.2, 0.4, 0.8, 1.6, 3.2, 6.4, 12.8, 25.6]$), was done to find the best combination of γ and C providing highest accuracy. All the features were standardized around mean zero and unit variance prior to grid search. To prevent possible overfitting with leave one out cross validation [28], the accuracy of the prediction was evaluated through a 10-fold cross validation scheme. The data was randomly divided in 10 equally sized subsets, of which 9 were used for training and the remaining one for testing. The mean accuracy after 10 repetitions was used to evaluate the prognostic capabilities of the selected variables.

The mean accuracy over 10 repetitions was calculated for all possible combinations of features to find the combination of features with the highest prediction accuracy.

Comparison between groups (benign vs malignant, and ER+HER2- vs ER+HER2+) for selected features was done using the Mann-Whitney test with Bonferroni correction.

The Mann Whitney test was performed in Matlab (2014a Mathworks, Natick), while SVM was implemented in Python programming language (version 3.6, Python Software Foundation, <https://www.python.org/>) using the scikit-learn library [29].

RESULTS

Figure 3 shows parametric maps from two patients with benign and malignant tumors, respectively.

Feature Selection

For all models, it was found that including more than 4 features for the SVM analysis failed to increase the 10-fold prediction accuracy, and in some cases decreased it (Figure 4).

Prediction of Benign vs Malignant tumors

Default accuracy by arbitrarily assigning all tumors as malignant is 0.55 (28/51) (red dashed line, Figure 4). Single feature accuracy from SVM with 10-fold cross-validation, as well as p-values obtained from univariate analysis are presented in Table 4.

For ADC model features, 10-fold accuracy of 0.92 was achieved by using either ADC_{mean} or ADC_{median} (Table 2). By combining two ADC features (multiple combinations possible, see Table 3) the accuracy increased to 0.94. The best accuracy (0.96) for ADC alone was obtained when combining 3 features (Table 2).

For IVIM model features, D_{mean} had the highest accuracy of 0.883, and the accuracy increased to 0.94 by combining either D_{mean} or D_{median} with one of D^*_{mean} , $D^*_{kurtosis}$, or f_{mean} . Highest accuracy for IVIM model was achieved by using 3 features: D_{mean} or D_{median} with D^*_{median} and $f_{kurtosis}$.

RED_{mean} alone achieved 0.96, and increasing the number of features failed to an increase accuracy.

When allowing combinations of features from all models simultaneously, an accuracy of 0.96 was achieved by combining 2 features (Table 2). The accuracy increased to 0.98 when combining 3 features. Finally, perfect predictive accuracy (1.0) was reached using 4 features combined. This combination included of RED_{median} , ADC_{std} , D^*_{std} with $ADC_{kurtosis}$, $D_{kurtosis}$ or D^*_{median} .

Prediction of ER+HER2- vs ER+HER2+ Tumors

Default accuracy by assigning all cases as ER+HER2- is 0.77 (20/26) (red dashed line, Figure 4). Single feature accuracy from SVM with 10-fold cross-validation, as well as p-values obtained from univariate analysis are presented in Table 5.

The most accurate single feature from the ADC model was $ADC_{kurtosis}$ with an accuracy of 0.8 (Table 3). The accuracy increased to 0.817 by combining $ADC_{kurtosis}$ with either ADC_{median} or ADC_{mean} .

For IVIM the single most accurate feature was D^*_{std} (0.767), and the accuracy increased to 0.8 using two features (multiple combinations possible, see Table 3). Again, a further increase in accuracy (0.817) was observed when using combinations of 3 features (D^*_{mean} , f_{mean} , and $f_{skewness}$).

RED_{std} alone gave an accuracy of 0.717. By combining it with $RED_{skewness}$ the accuracy increased to 0.8. By further adding RED_{median} into the combination, accuracy of 0.85 was achieved.

When allowing combinations of features from all models, an accuracy of 0.85 was achieved by several combinations of 2 features (see Table 3), while 4 features provided the best predictive power of 0.9 for ER+HER2- and ER+HER2+ (Table 35, multiple combinations possible).

DISCUSSION

There is already considerable evidence in the literature of the ability of ADC to differentiate between benign and malignant breast lesions [3, 5, 8, 10, 30]. Commonly, ROI mean or median value is the single feature used for classification. The results from

the current study show that adopting a machine learning approach including higher order statistical features from the ROI increases the classification accuracy of the ADC model. This indicates that the spatial distribution of ADC holds additional information on tumor heterogeneity relevant for the classification of malignancy, in agreement with previous studies [31]. Perfusion characteristics are known to be relevant for differentiation of malign and benign lesions [8, 9], and they improve accuracy compared to using only diffusion characteristics where only first order statistics are considered. In this work the IVIM model, which contains information about perfusion in addition to diffusion, achieved equivalent, but not higher accuracy compared with the ADC model when using 2 or more features. The Relative Enhanced Diffusivity model achieved its maximum accuracy for malignant vs benign classification when using only one feature; RED_{mean} . This result was equivalent to the best accuracy achieved for ADC and IVIM when using 3 or more features. Being sensitive to both microcirculation and diffusion effects, and more robust against noise than IVIM due to its simpler fitting approach RED contains sufficient information for the classification task. Interestingly, the high single-feature accuracy for the RED model was achieved without using any higher order statistical features associated with heterogeneity.

Finally, perfect differentiation of malignant and benign lesions in our cohort was achieved when using features from ADC, IVIM and RED models combined, indicating that these models contain complementary information that can be leveraged for clinical use (classifications). One possible interpretation is that in the combined approach, information about diffusion, perfusion and heterogeneity contribute together in an

optimized way. Furthermore, it is possible to obtain perfect classification using only lower order statistics from all models.

DWI-derived parameters are also able to predict the HER2 status of the malignant tumors, although with lower accuracy compared to the differentiation of benign and malignant lesions. Here, it is crucial to combine features in SVM to have predictive value, as none of the parameters itself is a significant differentiator in univariate analysis. In addition to mean/median among chosen features, skewness and kurtosis appear very important, emphasizing the importance of features associated with heterogeneity. Intratumor heterogeneity, which appear on any level (i.e. genes, cells, tissue, and clinical features) pose a huge challenge for diagnostics, and has implications for the further treatment selection [32].

Scan time and fitting complexity are both significantly higher for the IVIM model, using 13 b-values in this study, than for the ADC and RED models that are based on only three b-values. Our results show that classification accuracy for malignant versus benign is already very high using RED_{mean} alone, and the limited added clinical value of including the IVIM model must be balanced against the additional scan time and complexity of analysis. Optimal prediction of ER+HER2- and ER+HER2+ in this study can be achieved without IVIM features, and so accuracy would not be adversely affected using a simpler acquisition.

The classification accuracy for DWI-MRI parameters to differentiate benign and malignant tumors in our study were comparable to those reported in several other studies (91% and 86.7%) [3, 10]. Additionally, one of these also showed the increased accuracy when using RED over ADC shown in this study [10]. Contrary to other prior

studies [8, 9, 30], we observed no added value from consideration of IVIM over ADC. It may be because this work additionally considered ADC heterogeneity parameters, which increased the accuracy of multisided analysis of ADC.

For classification of HER2 status, varying results have been presented in literature. Significant differentiators in (non-Bonferroni adjusted) univariate analysis of DWI parameters have been found in other studies (not exclusively considering ER+ cases) including significantly lower mean D^* [15], higher mean ADC ($p=0.018$) [33], and a higher 90th percentile of D ($p=0.027$) [34] for HER2+ malignant tumors. It was reported that important parameters were mean, kurtosis and skewness of D^* for both (ER+HER2+ and ER+HER2-) as well as ADC kurtosis (for ER+HER2+ only) [15]. By combining skewness and mean of D^* , Cho, G.Y. et al. could distinguish ER+HER2- from all other cancer subtypes (AUC=0.8). As in our study it is important to combine first order with higher order statistical parameters to obtain optimal combination of diffusion parameters. However, other studies [35, 36] found neither correlation nor statistical differences between HER2 status and DWI-MRI parameters, as was the case in our univariate analysis.

Machine learning is common approach beyond univariate analysis, with many available algorithms. SVM was the choice in this work due to previous performance in breast cancer [18, 19]. In Cai et al, SVM (compared to k-nearest neighbors algorithm (KNN), Naïve Bayes classifier (NB), and logistic regression) achieved highest 10-fold accuracy for predicting malignancy status of tumor using ADC. By combining ADC, morphology, DCE kinetics and pathology, Cai et al found that SVM machine learning approach yielded accuracy of 92.4% [18]. Sutton et al. successfully implemented SVM using only

features developed from DCE images achieved overall accuracy of 71.2% for subtype groups of ERPR+, ERPR-HER2+ and TN (triple negative)[19].

It is important to note that all the differing accuracies provided by diffusion parameters, such as D and ADC, arise from the choice of b-value images from our dataset; a similar effect has been analyzed in previous studies for b-value choice [2, 37]. Standardization of acquisition strategies, diffusion models, and fitting strategies remains important for comparison across studies, although combination of parameter features across different models may ameliorate this problem. Additionally, IVIM-model is known to be noise sensitive, and several papers have explored the effect of different fitting algorithms [38, 39]. In this study, we applied the segmented fitting approach, whereas applying a Bayesian algorithm with a Gaussian or spatial prior [38] can be expected to return different histograms and thus potentially affect the machine learning outcome. The main limitation of this study is, however, the relatively small patient cohort. Importantly, the applied 10-fold validation scheme is less prone to overfitting compared to LOOCV [28]. A larger cohort would allow a separate test set and more rigorous validation. Thus, further validations in independent and larger cohorts are necessary to avoid overgeneralization.

In conclusion, the results from this study show that individual, conventional diffusion models and ROI statistics, including histogram moments, do not necessarily reflect all available information from clinical DW-MRI acquisitions. Our findings suggest that a SVM learning approach, using multiple features from a combination of diffusion models improves prediction accuracy for differentiation of benign and malignant breast lesions, and may further assist in differentiating HER2 status of ER+ lesions. The encouraging

predictive power of these combined features within this preliminary study demonstrates untapped potential of DW-MRI, and supports the inclusion of DW-MRI as part of a robust and in future potentially fully non-invasive, diagnostic process for breast cancer patients.

Acknowledgments:

We thank Dr. Mattijs Elschot for his valuable contribution regarding the machine learning algorithm.

REFERENCES

1. Le Bihan, D. and E. Breton, *Imagerie de diffusion in-vivo par résonance magnétique nucléaire*. Comptes-Rendus de l'Académie des Sciences, 1985. **93**(5): p. 27-34.
2. Cho, G.Y., L. Moy, J.L. Zhang, et al., *Comparison of fitting methods and b-value sampling strategies for intravoxel incoherent motion in breast cancer*. Magnetic Resonance in Medicine, 2015. **74**(4): p. 1077-1085.
3. Guo, Y., Y.-Q. Cai, Z.-L. Cai, et al., *Differentiation of clinically benign and malignant breast lesions using diffusion-weighted imaging*. Journal of Magnetic Resonance Imaging, 2002. **16**(2): p. 172-178.
4. Partridge, S.C., W.B. DeMartini, B.F. Kurland, P.R. Eby, S.W. White, and C.D. Lehman, *Differential diagnosis of mammographically and clinically occult breast lesions on diffusion-weighted MRI*. Journal of Magnetic Resonance Imaging, 2010. **31**(3): p. 562-570.
5. Rubesova, E., A.-S. Grell, V. De Maertelaer, T. Metens, S.-L. Chao, and M. Lemort, *Quantitative diffusion imaging in breast cancer: A clinical prospective study*. Journal of Magnetic Resonance Imaging, 2006. **24**(2): p. 319-324.
6. Koh, D.-M. and D.J. Collins, *Diffusion-Weighted MRI in the Body: Applications and Challenges in Oncology*. American Journal of Roentgenology, 2007. **188**(6): p. 1622-1635.
7. Le Bihan, D., E. Breton, D. Lallemand, M.L. Aubin, J. Vignaud, and M. Laval-Jeantet, *Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging*. Radiology, 1988. **168**(2): p. 497-505.
8. Sigmund, E.E., G.Y. Cho, S. Kim, et al., *Intravoxel incoherent motion imaging of tumor microenvironment in locally advanced breast cancer*. Magnetic Resonance in Medicine, 2011. **65**(5): p. 1437-1447.
9. Liu, C., C. Liang, Z. Liu, S. Zhang, and B. Huang, *Intravoxel incoherent motion (IVIM) in evaluation of breast lesions: Comparison with conventional DWI*. European Journal of Radiology, 2013. **82**(12): p. e782-e789.
10. Teruel, J.R., P.E. Goa, T.E. Sjobakk, A. Ostlie, H.E. Fjosne, and T.F. Bathen, *A Simplified Approach to Measure the Effect of the Microvasculature in Diffusion-weighted MR Imaging Applied to Breast Tumors: Preliminary Results*. Radiology, 2016. **281**(2): p. 373-381.
11. Jerome, N.P., M.R. Orton, J.A. d'Arcy, D.J. Collins, D.M. Koh, and M.O. Leach, *Comparison of free-breathing with navigator-controlled acquisition regimes in abdominal diffusion-weighted magnetic resonance images: Effect on ADC and IVIM statistics*. J Magn Reson Imaging, 2014. **39**(1): p. 235-40.
12. Miyazaki, K., N.P. Jerome, D.J. Collins, et al., *Demonstration of the reproducibility of free-breathing diffusion-weighted MRI and dynamic contrast enhanced MRI in children with solid tumours: a pilot study*. European Radiology, 2015. **25**(9): p. 2641-2650.
13. Jerome, N.P., K. Miyazaki, D.J. Collins, et al., *Repeatability of derived parameters from histograms following non-Gaussian diffusion modelling of diffusion-weighted imaging in a paediatric oncological cohort*. European Radiology, 2017. **27**(1): p. 345-353.
14. Zardavas, D., A. Irrthum, C. Swanton, and M. Piccart, *Clinical management of breast cancer heterogeneity*. Nat Rev Clin Oncol, 2015. **12**(7): p. 381-394.
15. Cho, G.Y., L. Moy, S.G. Kim, et al., *Evaluation of breast cancer using intravoxel incoherent motion (IVIM) histogram analysis: comparison with malignant status, histological subtype, and molecular prognostic factors*. Eur Radiol, 2016. **26**(8): p. 2547-58.
16. Perez, E.A., *Breast cancer management: opportunities and barriers to an individualized approach*. Oncologist, 2011. **16 Suppl 1**: p. 20-2.

17. Polyak, K., *Heterogeneity in breast cancer*. The Journal of Clinical Investigation, 2011. **121**(10): p. 3786-3788.
18. Cai, H., Y. Peng, C. Ou, M. Chen, and L. Li, *Diagnosis of breast masses from dynamic contrast-enhanced and diffusion-weighted MR: a machine learning approach*. PLoS One, 2014. **9**(1): p. e87387.
19. Sutton, E.J., B.Z. Dashevsky, J.H. Oh, et al., *Breast cancer molecular subtype classifier that incorporates MRI features*. J Magn Reson Imaging, 2016. **44**(1): p. 122-9.
20. Cortes, C. and V. Vapnik, *Support-vector networks*. Machine Learning, 1995. **20**(3): p. 273-297.
21. Polat, K. and S. Güneş, *Breast cancer diagnosis using least square support vector machine*. Digital Signal Processing, 2007. **17**(4): p. 694-701.
22. Hammond, M.E.H., D.F. Hayes, M. Dowsett, et al., *American Society of Clinical Oncology/College of American Pathologists Guideline Recommendations for Immunohistochemical Testing of Estrogen and Progesterone Receptors in Breast Cancer (Unabridged Version)*. Archives of Pathology & Laboratory Medicine, 2010. **134**(7): p. e48-e72.
23. Bartlett, J.M., J. Starczynski, N. Atkey, et al., *HER2 testing in the UK: recommendations for breast and gastric in-situ hybridisation methods*. J Clin Pathol, 2011. **64**(8): p. 649-53.
24. Teruel, J.R., H.E. Fjøsne, A. Østlie, et al., *Inhomogeneous static magnetic field-induced distortion correction applied to diffusion weighted MRI of the breast at 3T*. Magnetic Resonance in Medicine, 2015. **74**(4): p. 1138-1144.
25. Reese, T.G., O. Heid, R.M. Weisskoff, and V.J. Wedeen, *Reduction of eddy-current-induced distortion in diffusion MRI using a twice-refocused spin echo*. Magn Reson Med, 2003. **49**(1): p. 177-82.
26. Holland, D., J.M. Kuperman, and A.M. Dale, *Efficient correction of inhomogeneous static magnetic field-induced distortion in Echo Planar Imaging*. NeuroImage, 2010. **50**(1): p. 175-183.
27. Zitová, B. and J. Flusser, *Image registration methods: a survey*. Image and Vision Computing, 2003. **21**(11): p. 977-1000.
28. Rao, R.B., G. Fung, and R. Rosales. *On the dangers of cross-validation. An experimental evaluation*. in *Proceedings of the 2008 SIAM International Conference on Data Mining*. 2008. Society for Industrial and Applied Mathematics.
29. Pedregosa, F., Ga, #235, et al., *Scikit-learn: Machine Learning in Python*. J. Mach. Learn. Res., 2011. **12**: p. 2825-2830.
30. Bokacheva, L., J.B. Kaplan, D.D. Giri, et al., *Intravoxel incoherent motion diffusion-weighted MRI at 3.0 T differentiates malignant breast lesions from benign lesions and breast parenchyma*. Journal of Magnetic Resonance Imaging, 2014. **40**(4): p. 813-823.
31. Suo, S., K. Zhang, M. Cao, et al., *Characterization of breast masses as benign or malignant at 3.0T MRI with whole-lesion histogram analysis of the apparent diffusion coefficient*. Journal of Magnetic Resonance Imaging, 2016. **43**(4): p. 894-902.
32. Song, J.-L., C. Chen, J.-P. Yuan, and S.-R. Sun, *Progress in the clinical detection of heterogeneity in breast cancer*. Cancer Medicine, 2016. **5**(12): p. 3475-3488.
33. Jeh, S.K., S.H. Kim, H.S. Kim, et al., *Correlation of the apparent diffusion coefficient value and dynamic magnetic resonance imaging findings with prognostic factors in invasive ductal carcinoma*. J Magn Reson Imaging, 2011. **33**(1): p. 102-9.
34. Lee, Y.J., S.H. Kim, B.J. Kang, et al., *Intravoxel incoherent motion (IVIM)-derived parameters in diffusion-weighted MRI: Associations with prognostic factors in invasive ductal carcinoma*. Journal of Magnetic Resonance Imaging, 2017. **45**(5): p. 1394-1406.
35. Choi, S.Y., Y. Chang, H.J. Park, H.J. Kim, S.S. Hong, and D.Y. Seo, *Correlation of the apparent diffusion coefficient values on diffusion-weighted imaging with prognostic factors for breast cancer*. Br J Radiol, 2012. **85**(1016): p. e474-9.

36. Kamitani, T., Y. Matsuo, H. Yabuuchi, et al., *Correlations between apparent diffusion coefficient values and prognostic factors of breast cancer*. Magn Reson Med Sci, 2013. **12**(3): p. 193-9.
37. Lemke, A., B. Stieltjes, L.R. Schad, and F.B. Laun, *Toward an optimal distribution of b values for intravoxel incoherent motion imaging*. Magnetic Resonance Imaging, 2011. **29**(6): p. 766-776.
38. While, P.T., *A comparative simulation study of bayesian fitting approaches to intravoxel incoherent motion modeling in diffusion-weighted MRI*. Magnetic Resonance in Medicine, 2017: p. n/a-n/a.
39. Suo, S., N. Lin, H. Wang, et al., *Intravoxel incoherent motion diffusion-weighted MR imaging of breast cancer at 3.0 tesla: Comparison of different curve-fitting methods*. Journal of Magnetic Resonance Imaging, 2015. **42**(2): p. 362-370.

Table 1 Clinical characteristics of the patient cohort.

Characteristics	Result
Cancer	
No of cancers	28
Mean patient age (years)	53.7 (29.3 – 74.6)
Mean tumor volume (cm ³)	1.5 (0.2 – 4.25)
Histologic type	
Invasive ductal carcinoma	16
Invasive ductal carcinoma with ductal carcinoma in situ	8
Medullary carcinoma with ductal carcinoma in situ	1
Invasive lobular carcinoma	1
Mucinous carcinoma with ductal carcinoma in situ	1
Papillary carcinoma	1
Histologic grade	
1	6
2	9
3	9
2/3	2
Not analyzed	2
Receptor status	
ER+/PR+/HER2-	16
ER+/PR-/HER2-	4
ER+/PR+/HER2+	4
ER+/PR-/HER2+	2
ER-/PR-/HER2+	2
Mass/nonmass enhancement	
Mass	27
Nonmass	1
Benign	
No. of benign lesions	23
Mean patient age (years)	29.9 (20.6 – 53.3)
Mean tumor volume (cm ³)	8.1 (0.1 – 104.1)
Histologic type	
Fibroadenoma	14
Phyllodes	2
Fibroadenomatosis	1
Adenosis	1
No histologic analysis available	5

ER – estrogen receptor; *PR* – progesterone receptor; *HER2* – human epidermal growth factor

Table 2 Benign vs Malignant. Accuracy for SVM classification with 10-fold cross-validation, including the list of the feature combinations for each of the models with highest accuracy

No. of features	ADC	RED	IVIM	Combined
1	0.92	<u>0.96</u>	0.88	–
	Mean Median	Mean	D _{mean}	–
2	0.94	0.96	0.94	<u>0.963 (0.96[†])</u>
	Mean & skewness	Mean & median	D _{mean} & D* _{mean}	RED _{median} & ADC _{std}
	Mean & kurtosis	Mean & kurtosis	D _{mean} & D* _{kurtosis}	(other combinations [†])
	Std & median		D _{mean} & f _{mean}	
	Median & skewness		D _{median} & D* _{mean}	
	Median & kurtosis		D _{median} & D* _{kurtosis} D _{median} & f _{mean}	
3	0.96	0.96	0.96	<u>0.98</u>
	Mean, median & kurtosis	Mean, std & median	D _{mean} , D* _{median} & f _{kurtosis}	RED _{mean} , RED _{skewness} & D _{kurtosis}
	Mean, skewness & kurtosis		D _{median} , D* _{median} & f _{kurtosis}	RED _{median} , RED _{kurtosis} & ADC _{std}
	Std, median & kurtosis			RED _{median} , ADC _{mean} & ADC _{kurtosis} RED _{median} , ADC _{std} & D* _{std} RED _{median} , D _{mean} & D* _{skewness} RED _{median} , D _{median} & D* _{skewness}
4	0.96	0.94	0.96	<u>1</u>
	Mean, std, skewness & kurtosis	Mean, std, median & skewness	D _{mean} , D _{median} , D _{skewness} & D* _{median}	RED _{median} , ADC _{std} , ADC _{kurtosis} & D* _{std}
		Mean, std, median & kurtosis	D _{mean} , D _{median} , D* _{median} & f _{kurtosis}	RED _{median} , ADC _{std} , D _{kurtosis} & D* _{std}
		Mean, std, skewness & kurtosis	D _{std} , D _{median} , D* _{median} & f _{skewness}	RED _{median} , ADC _{std} , D* _{std} & D* _{median}
	Std, median, skewness & kurtosis	D _{median} , D _{skewness} , D* _{median} & f _{skewness}		

The highest accuracy for a specific number of features is underlined. ADC – apparent diffusion coefficient; RED – relative enhanced diffusivity; IVIM – intravoxel incoherent motion

† Complete list of combinations in supplementary

Table 3 ER+HER2- vs ER+HER2+: Accuracy for SVM classification with 10-fold cross-validation, including the list of the feature combinations for each of the models with highest accuracy

No. of features	ADC	RED	IVIM	Combined
1	<u>0.80</u> Kurtosis	0.72 std	0.77 D* _{std}	– –
2	0.817 mean & kurtosis median & kurtosis	0.80 std & skewness Mean & kurtosis	0.80 D _{std} & f _{skewness} D _{std} & f _{kurtosis} D _{kurtosis} & f _{std} D* _{std} & f _{skewness}	<u>0.85</u> RED _{skewness} & D _{kurtosis} ADC _{kurtosis} & D _{mean} ADC _{kurtosis} & D _{median}
3	0.817 Mean, median & kurtosis	<u>0.85</u> std, median & skewness	0.82 D* _{mean} , f _{mean} & f _{skewness}	0.85 Many combinations [†]
4	0.72 Mean, std, median & kurtosis	0.85 Mean, std, median & skewness	0.82 D _{skewness} , D* _{std} , D* _{median} & f _{std} D* _{std} , D* _{median} , f _{mean} & f _{skewness} D* _{kurtosis} , f _{mean} , f _{skewness} & f _{kurtosis}	<u>0.90</u> RED _{mean} , RED _{median} , RED _{skewness} & ADC _{mean} RED _{std} , RED _{skewness} , ADC _{mean} & ADC _{kurtosis} RED _{std} , RED _{skewness} , ADC _{median} & ADC _{kurtosis} RED _{median} , RED _{skewness} , ADC _{median} & D* _{kurtosis} RED _{skewness} , ADC _{mean} , ADC _{skewness} & ADC _{kurtosis} RED _{skewness} , ADC _{median} , ADC _{skewness} & ADC _{kurtosis}

The highest accuracy for a specific number of features is underlined. ADC – apparent diffusion coefficient; RED – relative enhanced diffusivity; IVIM – intravoxel incoherent motion

[†] Complete list of combinations in supplementary

Table 4 Benign vs Malignant. Single feature accuracy from SVM with 10-fold cross-validation, and p-values obtained from univariate Mann-Whitney testing of the histogram properties

<i>p-value (accuracy)</i>	<i>mean</i>	<i>std</i>	<i>median</i>	<i>skewness</i>	<i>kurtosis</i>
ADC	<0.00004 (0.92)	0.99 (0.59)	<0.00004 (0.92)	0.02 (0.74)	0.78 (0.50)
RED	<0.00004 (0.96)	<0.002 (0.77)	<0.00004 (0.94)	0.70 (0.63)	0.21 (0.34)
IVIM					
D	<0.00004 (0.88)	0.93 (0.51)	<0.00004 (0.86)	0.84 (0.47)	0.80 (0.42)
D*	0.04 (0.66)	0.06 (0.74)	0.04 (0.64)	0.05 (0.70)	0.044 (0.72)
f	<0.002 (0.72)	0.44 (0.55)	<0.002 (0.65)	0.53 (0.44)	0.83 (0.44)

Significant univariate differentiators are in bold font.

ADC – apparent diffusion coefficient; RED – relative enhanced diffusivity; IVIM – intravoxel incoherent motion; std – standard deviation

Bonferroni corrected significance threshold is $p=0.002$ (equivalent to the $p=0.05$ non-corrected), Bonferroni corrected high significance threshold is $p=0.00004$ (equivalent to the $p=0.001$ non-corrected).

Table 5 ER+HER2- vs ER+HER2-: Single feature accuracy from SVM with 10-fold cross-validation, and p-values obtained from univariate Mann-Whitney testing of the histogram properties

	<i>p-value (accuracy)</i>	<i>mean</i>	<i>std</i>	<i>median</i>	<i>skewness</i>	<i>kurtosis</i>
ADC		0.88 (0.75)	0.24 (0.70)	0.93 (0.72)	0.88 (0.70)	0.41 (0.80)
RED		0.60 (0.70)	0.29 (0.72)	0.48 (0.70)	0.60 (0.70)	0.02 (0.70)
IVIM						
D		0.78 (0.75)	0.21 (0.70)	0.88 (0.70)	0.45 (0.70)	0.65 (0.70)
D*		0.29 (0.77)	0.52 (0.70)	0.14 (0.70)	0.24 (0.70)	0.32 (0.70)
f		0.69 (0.70)	0.93 (0.70)	0.65 (0.70)	0.88 (0.70)	0.78 (0.70)

ADC – apparent diffusion coefficient; RED – relative enhanced diffusivity; IVIM – intravoxel incoherent motion; std – standard deviation

The Bonferroni corrected significance threshold is $p=0.002$ (equivalent to the $p=0.05$ non-corrected)

Figure Legends

Figure 1. Processing workflow diagram: **1.** Distortion correction (PA – posterior anterior, AP – anterior posterior direction), followed by motion correction when necessary, to obtain aligned images for all 13 b-values (in s/mm²); **2** 3D region of interest selection in high b-value image ; **3.** Voxel wise calculation of diffusion model parameters, ADC, RED, IVIM (D, f, D*); **4.** Univariate comparison (Mann-Whitney test) and/or SVM classification based on histogram properties of the diffusion parameters.

Figure 2. Diagram for SVM and feature selection: $k \in [1, 2, 3, 4, 5]$ for RED and ADC; $k \in [1, 2, \dots, 14, 15]$ for IVIM; $k \in [1, 2, \dots, 24, 25]$ when all possible features were used SVM – support vector machine; RBF – radial basis function kernel – nonlinear kernel described as $\exp(-\gamma \cdot \|x-x'\|^2)$; γ kernel size - determines the range of influence of samples selected by the model as support vectors; (C) regularization parameters - determines the number of selected support vectors, defining the complexity (smoothness) of the decision surface.

Figure 3. Parametric maps for the diffusion models overlaid the T2-weighted image (ADC mm, RED [%], D [mm²/s], f and D* [mm²/s]):
top row: Patient (Age 43 years) diagnosed with a benign tumor (adenosis)
bottom row: Patient (age 74 years) diagnosed with an invasive ductal carcinoma ((grade 1, ER+, PR-, HER2-)

Figure 4. Highest accuracy for each model (including combined) per number of features for both differentiation of benign vs malignant and for prediction of HER2 status of ER+. Red dashed line represents the default accuracy by arbitrarily assigning all tumors as malignant/ER+HER2-

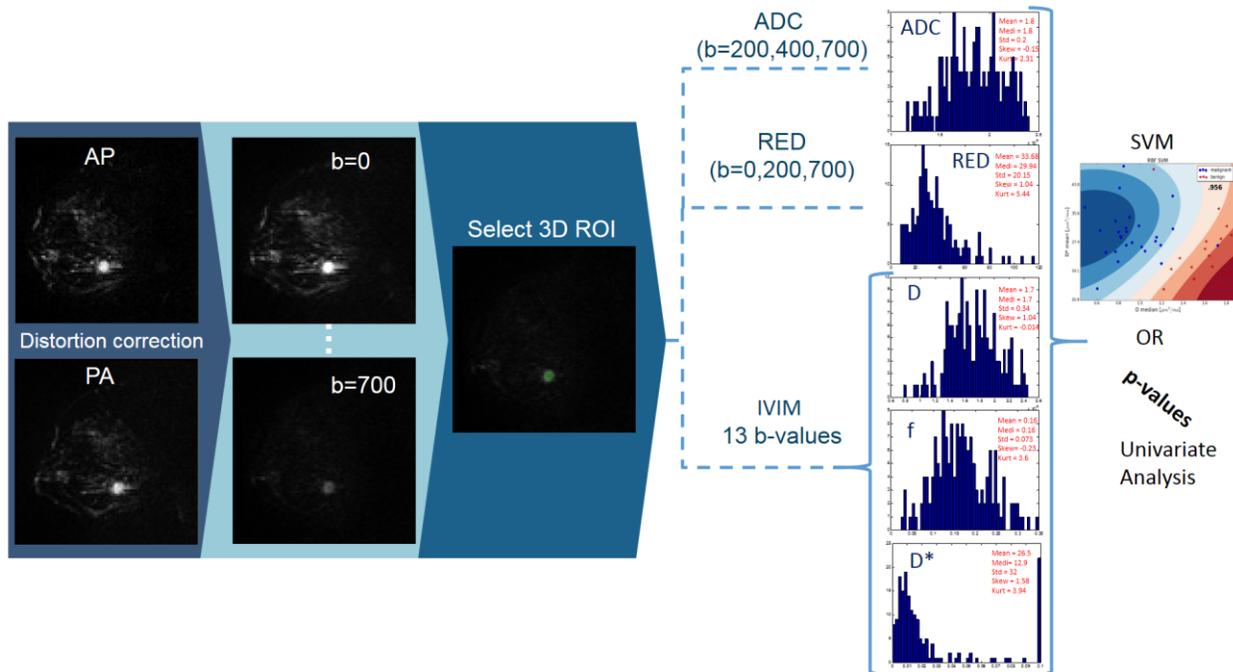


Figure 5. Processing workflow diagram: **1.** Distortion correction (PA – posterior anterior, AP – anterior posterior direction), followed by motion correction when necessary, to obtain aligned images for all 13 b -values (in s/mm²); **2** 3D region of interest selection in high b -value image ; **3.** Voxel wise calculation of diffusion model parameters, ADC, RED, IVIM (D , f , D^*); **4.** Univariate comparison (Mann-Whitney test) and/or SVM classification based on histogram properties of the diffusion parameters.

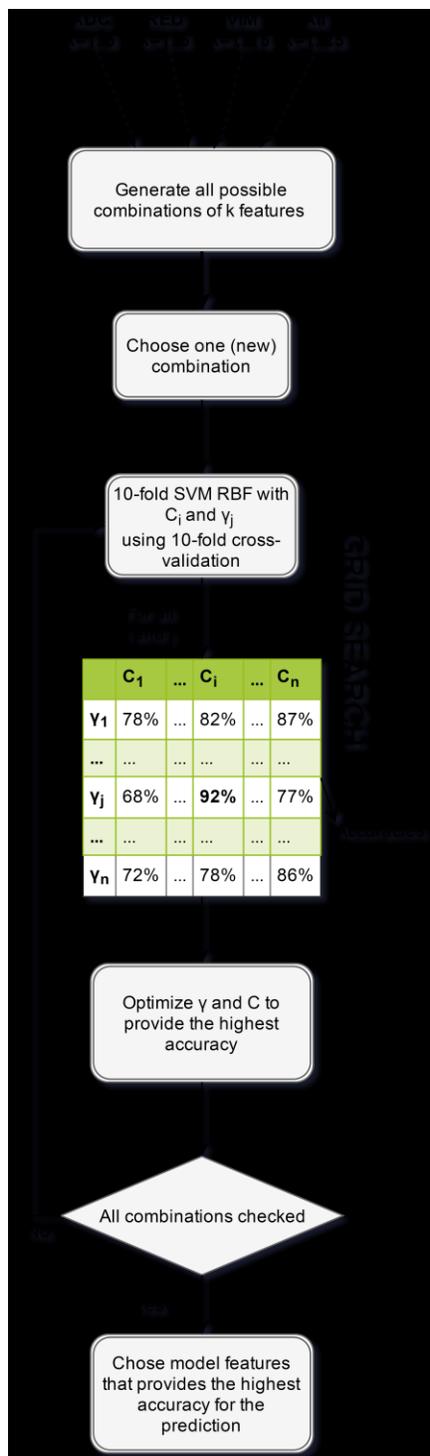


Figure 6. Diagram for SVM and feature selection: $k \in [1, 2, 3, 4, 5]$ for RED and ADC; $k \in [1, 2, \dots, 14, 15]$ for IVIM; $k \in [1, 2, \dots, 24, 25]$ when all possible features were used SVM – support vector machine; RBF – radial basis function kernel – nonlinear kernel described as $\exp(-\gamma \cdot \|x-x'\|^2)$; γ kernel size - determines the range of influence of samples selected by the model as support vectors; (C) regularization parameters - determines

the number of selected support vectors, defining the complexity (smoothness) of the decision surface.

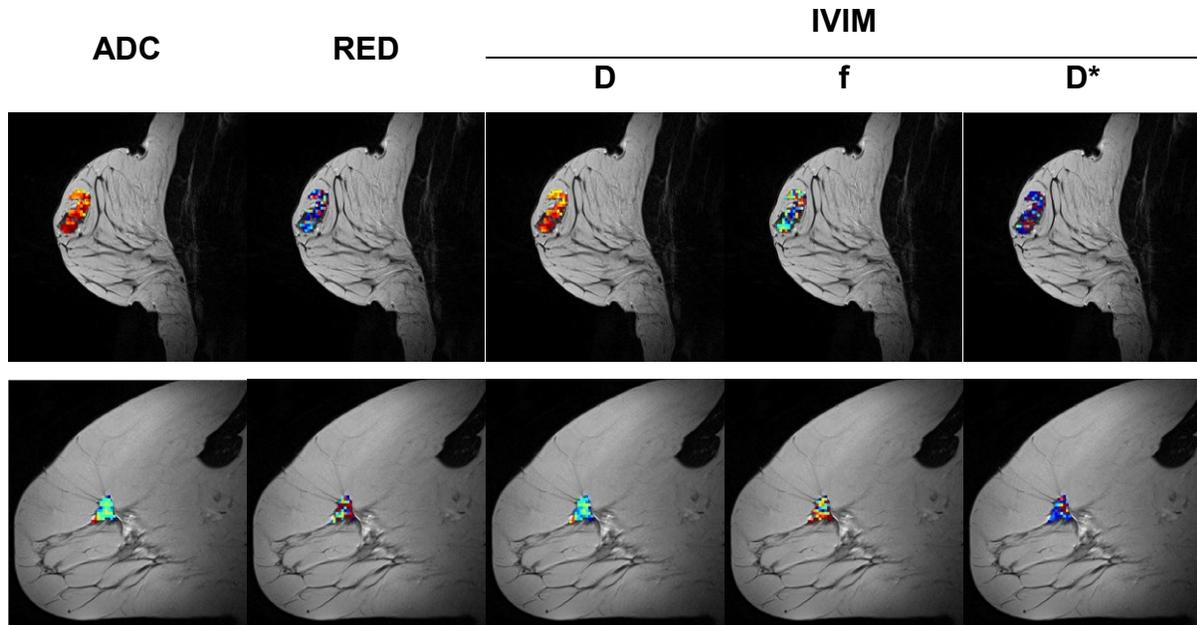


Figure 7. Parametric maps for the diffusion models overlaid the T2-weighted image (ADC mm, RED [%], D [mm^2/s], f and D* [mm^2/s]):
top row: Patient (Age 43 years) diagnosed with a benign tumor (adenosis)
bottom row: Patient (age 74 years) diagnosed with an invasive ductal carcinoma ((grade 1, ER+, PR-, HER2-)

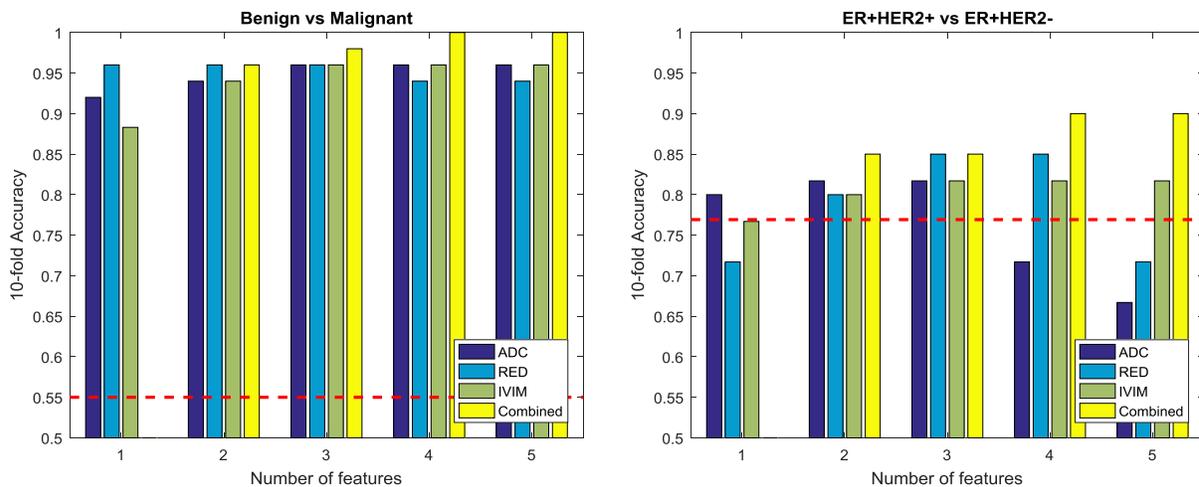


Figure 8. Highest accuracy for each model (including combined) per number of features for both differentiation of benign vs malignant and for prediction of HER2 status of ER+. Red dashed line represents the default accuracy by arbitrarily assigning all tumors as malignant/ER+HER2-