Accuracy of breast cancer lesion classification using IVIM DWI is improved by the inclusion of global or local prior knowledge with Bayesian methods

Igor Vidić[1†], Neil P. Jerome[2,3†], Tone F. Bathen[2], Pål E. Goa[1,3], Peter T. While[3*]

1 Department of Physics, NTNU – Norwegian University of Science and Technology, Trondheim, Norway
2 Department of Circulation and Medical Imaging, NTNU – Norwegian University of Science and Technology, Trondheim, Norway
3 Department of Radiology and Nuclear Medicine, St. Olav's University Hospital, Trondheim, Norway

† indicates equal contributions from joint first authors

***Correspondence:**
Peter T. While (Peter.Thomas.While@stolav.no)
Department of Radiology and Nuclear Medicine, St. Olav's University Hospital, Postbox 3250 Torgarden, NO-7006 Trondheim, Norway

Acknowledgements
We acknowledge support from the liaison Committee between the Central Norway Regional Health Authority and the Norwegian University of Science and Technology. We are grateful to Matthew R. Orton for useful insight on the Gaussian shrinkage prior.

Running Title

Bayesian IVIM improves breast lesion classification

Abstract

Background: Diffusion-weighted magnetic resonance imaging (DWI) has potential to non-invasively characterize breast cancer lesions; models such as intravoxel incoherent motion (IVIM) provide pseudodiffusion parameters that reflect tissue perfusion, but are dependent on details of acquisition and analysis strategy.

Purpose: To examine the effect of fitting algorithms, including conventional least-squares (LSQ) and segmented (SEG) methods as well as Bayesian methods with global shrinkage (BSP) and local spatial (FBM) priors, on the power of IVIM parameters to differentiate benign and malignant breast lesions.

Study Type: Prospective patient study.

Subjects: 61 patients with confirmed breast lesions.

Field Strength and Sequence: DWI (bipolar SE-EPI, 13 b-values) was included in a clinical MR protocol including T_2W and DCE-MRI on a 3T scanner.

Assessment: The IVIM model was fitted voxel-wise in lesion regions-of-interest (ROIs), and derived parameters were compared across methods within benign and malignant subgroups (correlation, coefficients of variation). Area-under receiver operator characteristic curves were calculated to determine discriminatory power of parameter combinations from all fitting methods.

Statistical Tests: Kruskal-Wallis, Mann-Whitney, Pearson correlation.

Results: All methods provided useful IVIM parameters; D was well-correlated across all methods ($r > 0.8$), with wider range for f and D^* (0.3–0.7). Fitting methods gave detectable differences in parameters, but all showed increased f and decreased D in malign lesions. D was the most discriminatory single parameter, with LSQ performing least well (AUC 0.83). In general, ROC AUCs were maximised by the inclusion of pseudodiffusion parameters, and by the use of Bayesian methods incorporating prior information (maximum AUC of 0.92 for BSP).

Conclusion: DWI performs well at classifying breast lesions, but careful consideration of analysis procedure can improve performance. D is the most discriminatory single parameter, but including pseudodiffusion parameters (f and D^*) increases ROC AUC. Bayesian methods outperformed conventional least-squares and segmented fitting methods for breast lesion classification.

Keywords:

Diffusion Imaging, Intravoxel Incoherent Motion, Breast Cancer, Classification, Bayesian

Accuracy of breast cancer lesion classification using IVIM DWI is improved by the inclusion of global or local prior knowledge with Bayesian methods

Introduction

Diffusion-Weighted Magnetic Resonance Imaging (DWI) is primarily sensitive to the Brownian motion of water molecules, which occupy the majority of the volume in all living tissue. In addition, DWI may also be sensitive to vascularity (perfusion), an effect first noticed by Le Bihan [1] and incorporated into the intravoxel incoherent motion (IVIM) model. Le Bihan postulated that water molecules involved in flow through randomly-oriented capillary vessels would effectively present like diffusion, only with a very high "pseudo-diffusion coefficient" D^*. The IVIM model is thus formulated as the sum of two exponential components, of empirical coefficients D^* and D, and the volume fraction f quantifying their relative contributions to the total signal. While IVIM has potential for non-invasive perfusion measurement, the method has not been widely clinically applied due to high noise sensitivity and lack of standardization, with studies demonstrating significantly different IVIM parameter values for the same tissue type [2] and in similar patient groups [3]; additionally, pseudodiffusion parameters f and D^* are generally less repeatable than D [4, 5]. Alongside data acquisition strategy, image quality, and physiological noise, IVIM parameters may depend to some degree on the signal fitting methods/algorithms used [2, 3, 6–12].

The most common method is to use some form of non-linear least squares (LSQ) fitting, which may be constrained or unconstrained to specified limits for each parameter; however, the noise sensitivity of this approach has led to exploration of alternatives that attempt to fit individual components separately in a segmented approach [3, 13-16], or to include some degree of prior knowledge, such as expectations of either local or global homogeneity, in more complex Bayesian-type approaches that are able to generate much cleaner parameter maps compared with those from non-linear least squares [17–21].

It is important for clinical interpretation of DWI in the breast, however, to note that it is by no means clear whether such Bayesian fitting methods result in model parameters more in agreement with underlying "true" values, and whether this necessarily results in more clinically useful results than more conventional methods. A recent simulation paper [19] considered this issue, and one finding was that these advanced Bayesian methods may suppress pseudodiffusion effects severely in the presence of measurement noise. This illustrates that care must be taken when applying Bayesian methods. The goal of Bayesian estimation is to frame results in the light of prior knowledge (or assumptions, known as the 'prior') about the true model parameters, and if the chosen prior is incorrect, the end result may be diverted away from the true values and give a less accurate result than a conventional method (i.e. LSQ). Thus, it is critical when applying Bayesian estimation to consider whether the chosen prior reflects the true behavior and properties of the tissue; to this end, additional and independent information about the tissue is required to act as reference and gold standard. Additionally, parameter estimates are typically derived from Bayesian methods by reporting specific metrics of the marginalized posterior distribution, such as the mean or mode, and the possible influence of the choice of metric on the final output needs to be explored [22].

Although useful for discussing and illuminating many issues, simulations cannot completely replace real experimental data in evaluating fitting methods, because the simulated data will ultimately be based on a model – and such models do not always capture all the relevant factors that influence real data, and will likely lack realistic spatial and statistical distributions. In this study, the utility of different IVIM fitting algorithms is assessed using the clinical classification of breast lesions, as either benign or malignant, as the success criterion. This explicitly avoids measures based on signal curve residuals, and focuses on the role of IVIM DWI in clinical decision making.

The specific hypotheses examined in the study are:

1. Bayesian fitting methods result in significantly different values for IVIM parameters in breast lesions compared to LSQ or segmented approaches.
2. Results returned from Bayesian fitting methods are sensitive to the choice of algorithm details – specifically, type of prior information and the choice of mean or mode from the obtained posterior distribution.
3. Bayesian fitting methods improve the classification of benign versus malignant voxels in breast lesions using the full IVIM parameter set, indicating better fit to the true perfusion properties.

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. Recruitment of patients was from October 2013 to August 2016.
Following MR examination, patients with malignant tumors underwent surgery and histopathologic analysis was performed on the resected mass. 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; 3 with significant 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. Clinical data is reported in Table 1. Data from this patient cohort was previously used in the study by Vidić

et al [14], and a subset (34 of 51) of the patients analyzed in this study were previously reported in a study by Teruel et al. describing the novel DWI metric relative enhanced diffusivity (RED) [23].

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, 700 s/mm^2 in either six (n=18, scan time 11 minutes) or three (n=33, scan time 6 minutes) directions. The protocol included one additional geometry-matched, non-diffusion-weighted (b=0s/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].

Dynamic contrast enhanced (DCE) and T_2-weighted images were also acquired. DCE scans consisted of 3D, T1-weighted, non-fat suppressed, gradient echo sequence, with TR 5.82ms, TE 2.18 ms, flip angle 15%, 256x256 matrix, in-plane resolution 0.7x0.7mm, slice thickness 2.5mm, and temporal resolution of 1 min, collected pre-contrast and at 7 consecutive time points following contrast administration. The DCE images were used for guidance of region of interest (ROI) selection in the DWI images. T_2-weighted images were non-fat suppressed, using a 2D turbo spin echo with TR 5500 ms, TE 118 ms, 256x256 matrix, in-plane resolution 0.7x0.7 mm, and slice thickness 2.5 mm.

Data Analysis

Preprocessing

All diffusion images were initially distortion-corrected via the method described by Holland et al [26], and validated in the breast by Teruel et al [24], using the phase-reversed b = 0 s.mm^{-2} images. A co-registration (3-dimensional, rigid) of all raw images to the corresponding b = 0 s.mm^{-2} image, using a normalized cross-correlation metric [27], was also performed in cases where there was substantial motion. Subsequently, trace images were calculated and entire lesions were segmented on the b = 700 s.mm^{-2} image guided by DCE images. Segmentation was performed by a basic scientist (I.V., 2 years of experience in breast imaging) with supervision by a breast radiologist (A.Ø., 20 years of experience). For each lesion, only the slice with the largest 2D ROI area was chosen for subsequent IVIM algorithm comparison.

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 Matlab routines (2016b Mathworks, Natick) developed in-house.

Fitting methods

The IVIM model is well-known, described by Le Bihan in the biexponential form:

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

where D is the tissue diffusion coefficient, f is the pseudodiffusion fraction, and D^* is the pseudodiffusion coefficient. Since IVIM is a model based on biophysical parameters, it is possible to impose boundaries on the parameters that are in line with expected underlying physiology; in this study, the following 'basic' limits were imposed: $f < 0.5$, implying that the pseudodiffusion component is not dominant in any voxel, $D < 0.003$ mm^{2}s^{-1}, the known limit for pure water diffusion at body temperature, and $D^* < 0.3$ mm^{2}s^{-1}, at which the D^* contribution is negligible at the lowest non-zero b-value in the study. Lower bounds were $f >$

0.0005, $D > 4.5\text{x}10^{-5}$ mm^2s^{-1}, and $D^* > 3.4\text{x}10^{-4}$ mm^2s^{-1} in line with previous work [17]. The following fitting methods were used:

1. Full non-linear least squares (LSQ)

In this conventional approach, all parameters are simultaneously estimated using a non-linear least-squares fitting function in Matlab (R2016b, Mathworks, Natick, MA, USA) and the trust-region-reflective algorithm, which enables imposition of constraints on the parameter space using the values given above. Initial values for the parameters were taken from literature on breast [11], in order to avoid bias that might be introduced by initialization using other methods (for example the segmented approach) where the cost function surface is likely to be flat.

2. Segmented approach (SEG)

This simplified fitting method has been applied to IVIM data in a number of studies [6, 11, 28], where the pseudodiffusion contribution to the signal is assumed negligible above a particular b-value, commonly 200 $s.mm^{-2}$. Fitting for monoexponential decay for $b \geq 200$ $s.mm^{-2}$ thus gives an estimation of D directly; the zero intercept (S_{int}) of this monoexponential decay is used to estimate the pseudodiffusion fraction parameter f:

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

Finally, D^* is estimated by a constrained LSQ fitting to the IVIM equation as above, but with fixed values for D and f.

3. Bayesian uniform prior (BUP)

The framework of Bayesian estimation is to include prior information when conducting a fitting operation; this prior can take various forms, and may be strong or weak in relation to the information contained in the data itself, and can be visualized by comparing the sharpness

of the peak in the likelihood function (negative log of the least-squares cost function for Gaussian errors), with the peak (if any) in the prior distribution function. The simplest type of prior is known as ‘minimally-informative’, which implies that estimation is driven purely by the data, and not by the prior distribution. This method uses the Bayesian framework to produce a posterior distribution of parameter values from which final estimates can be obtained from central tendency measures, such as the mean or the mode. When a uniform prior distribution is used, the mode of the posterior distribution is the same as the maximum likelihood estimate (which for Gaussian errors will be the same as the optimal least-squares estimate). Application of the Bayesian approach with minimally informative priors for IVIM data was first presented by Neil and Bretthorst [29].

4. Bayesian Gaussian “shrinkage” prior (BSP)

An alternative to using a uniform prior is to use a Gaussian prior with a mean and variance that reflect the expected location and spread in the values being estimated. However, providing fixed prior expected values for IVIM parameters is often unrealistic in real-world cases as these may not be known in advance. Another approach is to use the data in a collection of voxels (i.e. a lesion) to estimate characteristics of the Gaussian prior (mean and standard deviation) in addition to the IVIM parameters in each voxel. This prior model assumes a single tissue class, which for sufficiently homogeneous lesions will be a reasonable approximation in practice. This approach proposed by Orton et al [17] uses an LSQ fit to initialize a Markov Chain Monte Carlo (MCMC) estimation algorithm which then generates estimates of the Gaussian prior parameters and IVIM voxel parameters at every step. This form of prior distribution is often referred to as a ‘shrinkage prior’ because, for voxels that have both high IVIM parameter uncertainty and have estimates that are distant from the majority of the other voxels, there is a tendency for the estimates to ‘shrink’ towards the values estimated by the majority. High uncertainty can be due to poor SNR in the given voxel, or because the model is ill-conditioned (e.g. bi-exponential rate constants that are not

well separated). A more detailed description of the shrinkage prior approach along with pseudo code for implementation is given in Orton et al [17]. Specific for this work, whole-lesion ROIs were used to provide a large enough number of voxels for estimation of the initial Gaussian prior distribution, but only the largest slice was used in the analysis of the final IVIM parameters and comparison to other methods.

Both Bayesian fittings, BUP and BSP, used the Markov Chain Monte Carlo algorithm, with 2000 burn-in steps and 20,000 chain steps.

5. Spatial homogeneity prior – fusion bootstrap moves (FBM)

Unlike BSP, which considers all voxels form a single distribution, the spatial homogeneity prior from Freiman et al [18] assumes a level of voxel-to-neighboring-voxel similarity; the amount of homogeneity between neighboring voxels is governed by an additional parameter α in the energy minimization, which balances between returning a completely smooth map and a spatially-naïve map (as given by LSQ, which treats voxels entirely independently). Thus α is a case-specific parameter, and values were assessed and optimized visually. Additionally, the relative degree of smoothing between parameters is controlled by a weighting vector, W, and in this work the effective weighting of each parameter was approximately normalized by setting the three elements of W equal to the inverses of the literature values [11] that were also used to initialize LSQ.

The LSQ estimates provide the initial maps, and new proposals are generated using bootstrap resampling of the data. Iterations are conducted using 'fusion moves', which combine the current and subsequent solutions at each step to minimize the corresponding energy function. Further details are provided by While [4].

Statistical Analysis

Statistical analysis between fitting strategies was performed using the parameters derived from IVIM fitting within the 2D ROI, defined for each lesion as the slice with the largest area. The Kruskal-Wallis test was used to evaluate if parameter estimates came from the same underlying distribution. This was followed post-hoc by a Mann-Whitney test for multiple comparisons with Bonferroni correction. Receiver operating characteristic (ROC) along with area under the curves (AUC) were used for assessing differentiation ability of the parameters calculated from each method. Coefficients of variance, and Pearson correlation coefficients, were calculated for all parameters. All the statistical analysis was performed in Matlab (2016b Mathworks, Natick).

Results

Representative IVIM parameter maps are shown in Figure 1 for two tumors, with the corresponding value box plots; the variation in the parameters conforms roughly to expectation, with D being more consistent across the lesion and between fitting methods compared to f and D^* being increasingly variable in both contexts. Visually, the D maps are similar across all methods, whereas the non-Bayesian methods (SEG, LSQ) show similar voxels of increased pseudodiffusion fraction f. The BUP_{mode} maps follow the LSQ closely, which is unsurprising from theory. Notable is that the D^* map using the local spatial homogeneity assumption in FBM appears much smoother, giving it a cleaner appearance (while not necessarily changing the lesion summary statistics); from visual inspection to avoid excessive smoothing, a value of $\alpha = 4$ for the FBM method was used for all cases in this study.

Histograms showing the distribution of IVIM parameters in all voxels across the cohort, separated by their pathological benign/malignant classification, are shown in Figure 2. The distributions in general replicate the known characteristics of breast lesions, specifically an increase in f for malignant lesions and the corresponding decrease in D, interpreted as meaning increased vascularization and cellularity, respectively. For some fitting methods,

most notably LSQ and SEG, distributions for D^* and f contain a fraction of voxels at the upper limit, which suggests a failure in fitting that may arise from extreme values of underlying characteristics, or from low-quality data in these voxels.

The parameter distributions themselves vary across fitting methods to some degree (see Figure 3 for Mann-Whitney test results), and the Pearson correlations between parameters across fitting methods are given in Table 2. D appears to be consistently estimated across methods and has high inter-method correlations, whereas a wide range of correlation coefficients are seen for f and D^* - from above 0.9 when comparing mean and mode of BSP (f and D^*) to less than 0.4 when comparing BSP with LSQ (f) and SEG (D^*), and for FBM with LSQ (f and D^*), SEG (D^*) and BUP (D^*).

A proxy for parameter precision was measured by the coefficient of variation (CV) for each parameter across the lesion, shown in Table 3. Since this necessarily contains underlying real variation it is not a true measure of precision, but relative values are illustrative; namely, D appears the most precise, followed by f and then D^*; the pseudodiffusion parameters exhibit lower CVs in malign lesions. In general, FBM and both BSP approaches appear the most precise, although this is an expected consequence of these fitting approaches.

It is important to note that since neither the underlying 'true' values nor the true applicability of the model are known, the most useful assessment of these resulting distributions is how they inform clinical decision making, in this case discrimination of malignant and benign tumors. The associated ROC curves, and the resulting AUC values, for the IVIM fitting methods are shown in Figure 4, with individual parameters in 4a) and pair combinations from logistic regression in 4b). Again, the overall picture is in line with previous work, with D performing well (minimum AUC 0.83 from LSQ, maximum 0.9 from both BSP metrics), and pseudodiffusion parameters f and D^* performing poorly in isolation (maximum AUCs: 0.74 and 0.65, for f with BSP_{mode} and D^* with FBM). In practice, of course, all IVIM parameters arise from the same scanning sequence, and so are not derived separately; ROC combinations of increasing complexity, from D, to $(D+f)$ or $(D+D^*)$, and to $(D+f+D^*)$, illustrated in figure

4c), show an increase in discriminatory power. Overall, the best classification accuracies are provided by methods that go beyond simple curve fitting and include spatial prior knowledge, with the Bayesian shrinkage prior (BSP, and taking the posterior mean) and fusion bootstrap moves (FBM) providing AUCs of 0.92 and 0.9, respectively. Of particular note is that the conventional least-squares approach to fitting, often the default, gives the lowest AUC in all comparisons, and indicates that the segmented approach should be a preferred default method in the absence of facility to employ the more complex fitting algorithms.

Discussion

Firstly, it should be recognized that all fitting methods were 'successful' in the simple sense of returning IVIM parameters, notionally providing superficially similar information about the breast lesions. Closer examination of the effects of the fitting detail, however, reveals an influence that can directly affect the conclusions drawn in terms of disease characterization and, ultimately, patient treatment. While DWI is rarely the only MR modality used for diagnosis of breast cancer, that is only the current status and it is critical that exploration of possibilities and limits be performed, such that coherent strategies are developed and standardized within the DWI community. Only then can we collectively maximize the utility of and trust in the additional information that DWI can provide in the context of an evolving understanding of disease presentation and management. Acknowledging all of the factors that influence the IVIM parameter estimates, including those which determine the behavior of the signal itself [30,31], as well as the choices made for any given analysis for a particular data set, is therefore essential for useful interpretation of results.

Examining the hypotheses that were tested in this study, the results clearly indicate that the choice of fitting method influences the resulting IVIM parameter estimates, most evident in, but not restricted to, the pseudodiffusion parameters f and D^*. A clear effect of using a Bayesian shrinkage or local spatial homogeneity prior is to reduce the variation in f and D^*, including from extreme values. Spatial priors therefore give visually smoother parameter

maps, which is perhaps more relevant to the use of DWI for lesion depiction and localization, rather than quantitative functional analysis, but this may nevertheless contribute to increased acceptance of IVIM as a clinical tool.

The results from this study strongly suggest that attention to the details of diffusion imaging fitting method and execution results in better decisions, as measured by accuracy of breast lesion classification. An increase in ROC AUC was found for moving from simple non-linear least-squares fitting, unfortunately both the worst-performing and a common default method, to more advanced methods; this increase in AUC ranged in value, both when considering D in isolation (from +0.03 for BUP_{mode}, to +0.07 for BSP methods), and when combining all parameters (from +0.01 for BUP_{mode}, to +0.06 for BSP_{mean}). It is interesting to note that the segmented approach outperformed the least-squares method; a simplified version of this approach involves estimation of only *D* and *f*, and can thus be performed using only 3 b-values in a reduced scan time [32–36]. A relevant question is therefore whether *D** contributes sufficient added clinical value to justify the added complexity in both acquisition and analysis, given the performance of simpler approaches. We observe that *D** suffers most variation across lesions and fitting methods, and has relatively low classification accuracy in isolation (ROC AUC 0.65), but note that when added to a regression combination of all parameters *D** does increase accuracy, regardless of fitting method, actually adding more value than *f* when added to *D* (although *f* is always provided alongside *D**). The additional accuracy is small, however, so it is likely to remain a subjective choice whether the gain is worth the increased scanning and analysis demand. Nevertheless, this study contributes to the evidence in the literature that the IVIM effect is present and informative in breast lesions, albeit best pursued with the additional leverage provided by Bayesian approaches.

It has been found by previous studies that the choice of Bayesian methods for fitting IVIM data can give lower variability, especially for *f* and *D** [17, 18, 29], but runs the potential risk of obscuring real features in regions where there is a high uncertainty associated with the parameters [19]. Similarly, when the choice of posterior summary statistic can influence the

resulting estimations [22], it is critical that analysis is not performed blind to these details. Additionally, Bayesian methods require substantially longer for analysis, and some studies argue that simpler analysis can provide similar results [10, 34]. However comparing across studies with different cohorts and acquisition protocols is problematic, and these studies do not go as far as examining effects on classification accuracy. While Bayesian methods are substantially more computationally expensive, which makes implementation more cumbersome, this is decreasingly a genuine barrier given the availability and cost of computational power.

One limitation of this study is that the effect of parameter boundaries is not explored; it is possible to impose tighter boundaries to better reflect expected values, although this raises the issue of how to treat/interpret voxels at these boundaries. In this study, liberal boundaries were chosen, to avoid switching of D and D^* compartments (i.e. $f > 0.5$) and to remain within constraints of physical limits. In the case where Bayesian methods are recommended, however, the effect of imposed parameter boundaries is minimal. Other algorithms have associated specific limitations; in the case of the Gaussian shrinkage prior, there is an implicit assumption that there is a single tissue class. Thus, analysis on a defined ROI works well if this assumption is valid, but the method cannot be applied to whole images, and this is a limitation not shared by the other methods. This means only that IVIM parameter maps from BSP cannot be used for lesion delineation, though this is most commonly done on DCE or high-b-value images. One limitation of the FBM approach in this study is that the degree of local homogeneity is governed by a parameter (α) that must be empirically selected, and this may be expected to be tissue and/or disease specific. A strategy for determining the optimal value for α would be a valuable addition to the FBM method. One final limitation of the study is the cohort data from a single center, and thus the conclusions are not implicitly generalized for data acquired across locations and scanners.

Ultimately, the driving force behind DWI acquisition and analysis is creating the best information for patient care, and while neither DWI nor IVIM are new ideas, there remains much discussion – but no consensus - on optimal execution. A meta-analysis from 2008 of 44 different studies found that, for lesion classification, DCE had a pooled weighted sensitivity of 0.90 and specificity of 0.72 [37], and so although DWI is often considered an adjunct modality rather than an explicit alternative, the results in this study clearly show a competitive accuracy derived from more careful DWI processing. Explicitly framing the role of IVIM parameters in terms of ROC accuracy for classification, this study provides increased support for the use of more advanced, Bayesian fitting algorithms, such as Gaussian shrinkage and local spatial homogeneity priors, and for the inclusion of pseudodiffusion parameters in lesion classification.

References

1. Le Bihan D, Breton E, Lallemand D, et al. (1988) Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. Radiology 168:497–505.
2. Barbieri S, Donati OF, Froehlich JM, Thoeny HC (2016) Impact of the calculation algorithm on biexponential fitting of diffusion-weighted MRI in upper abdominal organs. Magn Reson Med 75:2175–2184.
3. Suo S, Lin N, Wang H, et al. (2015) Intravoxel incoherent motion diffusion-weighted MR imaging of breast cancer at 3.0 tesla: Comparison of different curve-fitting methods. J Magn Reson Imaging 42:362–370.
4. Orton MR, Jerome NP, Rata M, Koh D-M (2018) IVIM in the Body: A General Overview. In: Le Bihan D, Iima M, Federau C, Sigmund EE (eds) Intravoxel Incoherent Motion MRI Princ. Appl. Pan Stanford Publishing Pte. Ltd., pp 145–174
5. Jerome NP, Miyazaki K, Collins DJ, et al. (2017) Repeatability of derived parameters from histograms following non-Gaussian diffusion modelling of diffusion-weighted imaging in a paediatric oncological cohort. Eur Radiol 27:345–353.
6. Meeus EM, Novak J, Withey SB, et al. (2017) Evaluation of Intravoxel Incoherent Motion Fitting Methods in Low-Perfused Tissue. J Magn, Reson Imag 45:1325–1334.
7. Merisaari H, Movahedi P, Perez IM, et al. (2017) Fitting methods for intravoxel incoherent motion imaging of prostate cancer on region of interest level: Repeatability and gleason score prediction. Magn Reson Med 77:1249–1264.
8. Keil VC, Mädler B, Gielen GH, et al. (2017) Intravoxel incoherent motion MRI in the brain: Impact of the fitting model on perfusion fraction and lesion differentiability. J Magn Reson Imaging 46:1187–1199.

9. Park HJ, Sung YS, Lee SS, et al. (2017) Intravoxel incoherent motion diffusion-weighted MRI of the abdomen: The effect of fitting algorithms on the accuracy and reliability of the parameters. J Magn Reson Imaging 45:1637–1647.

10. Gurney-Champion OJ, Klaassen R, Froeling M, et al. (2018) Comparison of six fit algorithms for the intravoxel incoherent motion model of diffusionweighted magnetic resonance imaging data of pancreatic cancer patients. PLoS One 13:1–18.

11. Cho GY, Moy L, Zhang JL, et al. (2015) Comparison of fitting methods and b-value sampling strategies for intravoxel incoherent motion in breast cancer. Magn Reson Med 74:1077–1085.

12. While PT (2018) Advanced Methods for IVIM Parameter Estimation. In: Le Bihan D, Iima M, Federau C, Sigmund EE (eds) Intravoxel Incoherent Motion MRI Princ. Appl. Pan Stanford Publishing Pte. Ltd., pp 449–484

13. Sigmund EE, Cho GY, Kim S, et al. (2011) Intravoxel incoherent motion imaging of tumor microenvironment in locally advanced breast cancer. Magn Reson Med 65:1437–1447.

14. Vidic I, Egnell L, Jerome NP, et al. (2018) Support Vector Machine for Breast Cancer Classification Using Diffusion-Weighted MRI Histogram Features : Preliminary Study. J Magn Reson Imaging 47:1205–1216.

15. Bokacheva L, Kaplan JB, Giri DD, Patil S, Gnanasigamani M, Nyman CG, Deasy JO, Morris EA, Thakur SB (2014) Intravoxel incoherent motion diffusion-weighted MRI at 3.0 T differentiates malignant breast lesions from benign lesions and breast parenchyma. *J. Magn. Reson. Imaging* 40: 813–823

16. Chen F, Chen P, Muhammed HH, Zhang J (2017) Intravoxel Incoherent Motion Diffusion for Identification of Breast Malignant and Benign Tumors Using Chemometrics. *Biomed Res. Int.* 2017: 3845409

17. Orton MR, Collins DJ, Koh D, Leach MO (2014) Improved Intravoxel Incoherent Motion Analysis of DiffusionWeighted Imaging by Data Driven Bayesian Modeling. Magn Reson Med 71:411–420.

18. Freiman M, Perez-Rossello JM, Callahan MJ, et al. (2013) Reliable estimation of Incoherent Motion parametric maps from diffusion-weighted MRI using fusion bootstrap moves. Med Image Anal 17:325–36.

19. While PT (2017) A Comparative Simulation Study of Bayesian Fitting Approaches to Intravoxel Incoherent Motion Modeling in Diffusion-Weighted MRI. Magn Reson Med 78:2373–2387.

20. Taimouri V, Afacan O, Perez-Rossello JM, et al. (2015) Spatially constrained incoherent motion method improves diffusion-weighted MRI signal decay analysis in the liver and spleen. Med Phys 42:1895–1903.

21. Spinner GR, Von Deuster C, Tezcan KC, et al. (2017) Bayesian intravoxel incoherent motion parameter mapping in the human heart. J Cardiovasc Magn Reson 19:1–14.

22. Gustafsson O, Montelius M, Starck G, Ljungberg M (2018) Impact of prior distributions and central tendency measures on Bayesian intravoxel incoherent motion model fitting. Magn Reson Med 79:1674–1683.

23. Teruel JR, Goa PE, Sjøbakk TE, et al. (2016) A Simplified Approach to Measure the Effect of the Microvasculature in Diffusion-weighted MR Imaging Applied to Breast Tumors: Preliminary Results. Radiology 281:373–381.

24. Teruel JR, Fjøsne HE, Østlie A, et al. (2015) Inhomogeneous static magnetic field-induced distortion correction applied to diffusion weighted MRI of the breast at 3T. Magn Reson Med 74:1138–1144.

25. Alexander AL, Tsuruda JS, Parker DL (1997) Elimination of eddy current artifacts in diffusion-weighted echo-planar images: The use of bipolar gradients. Magn Reson Med 38:1016–1021.

26. Holland D, Kuperman JM, Dale AM (2010) Efficient correction of inhomogeneous static magnetic field-induced distortion in Echo Planar Imaging. Neuroimage 50:175–183.

27. Flusser J, Zitova B (2003) Image registration methods : a survey. Image Vis Comput 21:977–1000.

28. Patel J, Sigmund EE, Rusinek H, et al. (2010) Diagnosis of cirrhosis with intravoxel incoherent motion diffusion MRI and dynamic contrast-enhanced MRI alone and in combination: Preliminary experience. J Magn Reson Imaging 31:589–600.

29. Neil JJ, Bretthorst GL (1993) On the Use of Bayesian Probability Theory for Analysis of Exponential Decay Data: An Example Taken from Intravoxel Incoherent Motion Experiments. Magn Reson Med 29:642–647.

30. Jerome NP, D'Arcy JA, Feiweier T, et al. (2016) Extended T2-IVIM model for correction of TE dependence of pseudo-diffusion volume fraction in clinical diffusion-weighted magnetic resonance imaging. Phys Med Biol 61:N667–N680.

31. Iima M, Nobashi T, Imai H, et al. (2018) Effects of diffusion time on non-Gaussian diffusion and intravoxel incoherent motion (IVIM) MRI parameters in breast cancer and hepatocellular carcinoma xenograft models. Acta Radiol Open 7:205846011775156.

32. Penner AH, Sprinkart AM, Kukuk GM, et al. (2013) Intravoxel incoherent motion model-based liver lesion characterisation from three b-value diffusion-weighted MRI. Eur Radiol 23:2773–2783.

33. Pieper CC, Willinek WA, Meyer C, et al. (2016) Intravoxel Incoherent Motion Diffusion-Weighted MR Imaging for Prediction of Early Arterial Blood Flow Stasis in Radioembolization of Breast Cancer Liver Metastases. J Vasc Interv Radiol 27:1320–1328.

34. Jalnefjord O, Andersson M, Montelius M, et al. (2018) Comparison of methods for estimation of the intravoxel incoherent motion (IVIM) diffusion coefficient (D) and perfusion fraction (f). Magn Reson Mater Physics, Biol Med 31:715–723.

35. Bihan D Le, Breton E, Lallemand D, et al. (1988) Perfusion in intravoxel incoherent motion mr imaging. Radiology 168:497–505.

36. While PT, Teruel JR, Vidić I, et al. (2018) Relative enhanced diffusivity : noise sensitivity , protocol optimization , and the relation to intravoxel incoherent motion. Magn Reson Mater Phy 31:425–438.

37. Peters NHGM, Borel Rinkes IHM, Zuithoff NPA, et al. (2008) Meta-Analysis of MR Imaging in the Diagnosis of Breast Lesions. Radiology 246:116–124.

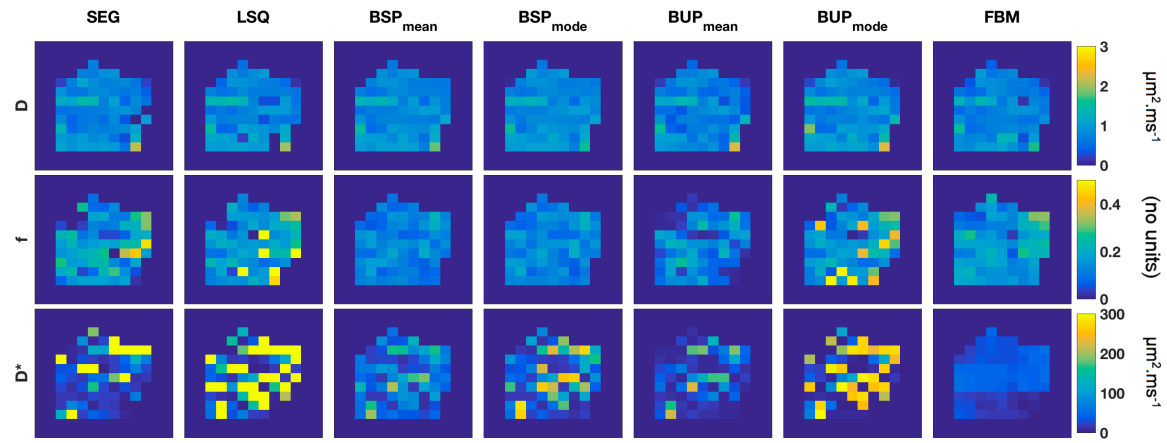

Figure 1a. IVIM parameter maps from each fitting method, taken from two typical lesions (malignant and benign in figures 1a and 1b, respectively). Fitting using prior information either across the lesion (BSP) or locally (FBM) yields smoother maps for both *f* and *D**. BSP mean and mode are generally similar, as expected, as are BUP_{mode} and LSQ, whereas BUP_{mean} has some apparent smoothing. In general, *D* maps are clearest, and *D** show the most variable characteristics. Note that voxel size appears different due to image cropping.

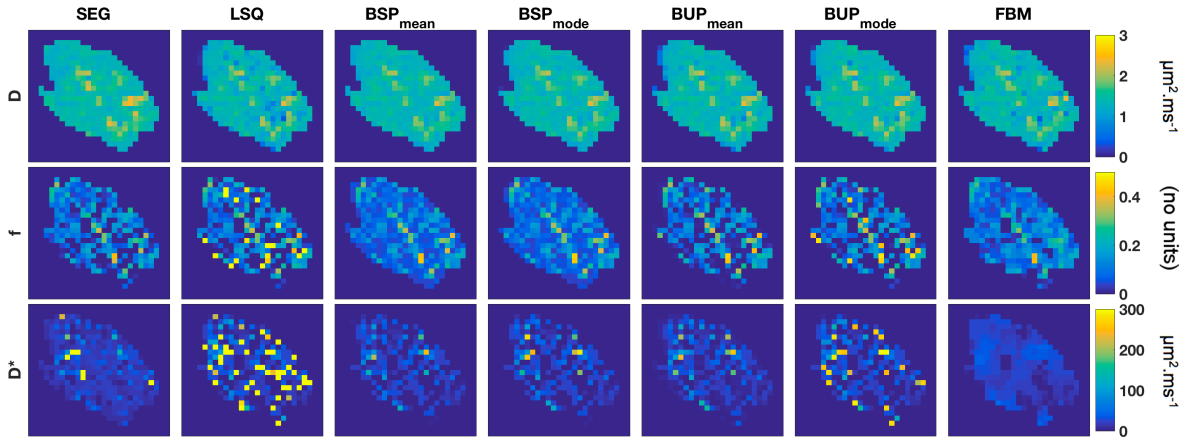

Figure 1b. IVIM parameter maps from each fitting method, taken from two typical lesions (malignant and benign in figures 1a and 1b, respectively). Fitting using prior information either across the lesion (BSP) or locally (FBM) yields smoother maps for both *f* and *D**. BSP mean and mode are generally similar, as expected, as are BUP_{mode} and LSQ, whereas BUP_{mean} has some apparent smoothing. In general, *D* maps are clearest, and *D** show the most variable characteristics. Note that voxel size appears different due to image cropping.

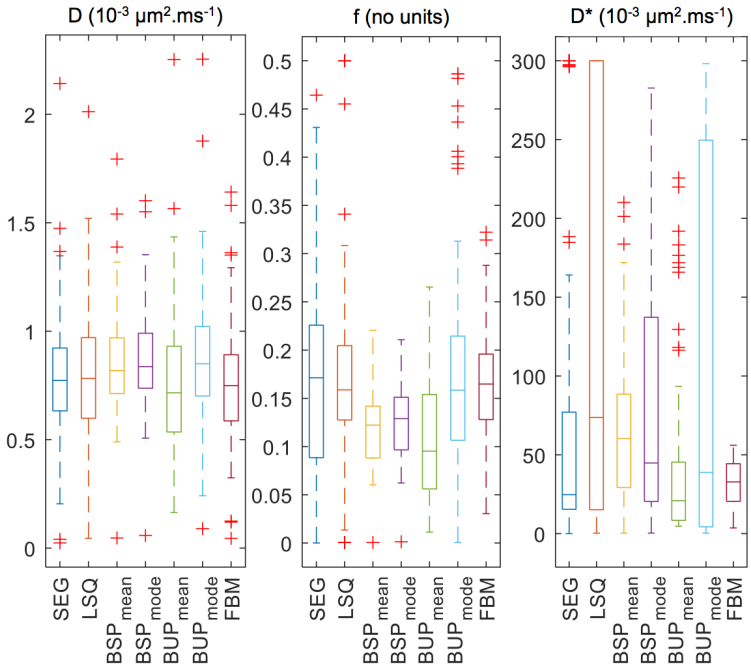

Figure 1c. Corresponding box plots for the lesion depicted in a). The more consistent behavior seen for *D* across the methods is not matched for either *f* or *D**, where resulting values are more dependent on the method employed.

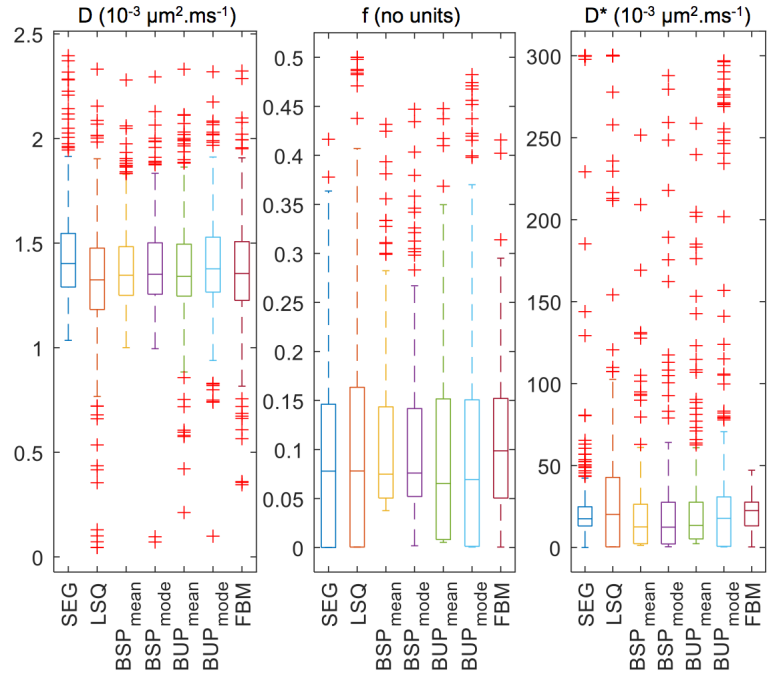

Figure 1d. Corresponding box plots for the lesion depicted in b). The more consistent behavior seen for *D* across the methods is not matched for either *f* or *D**, where resulting values are more dependent on the method employed.

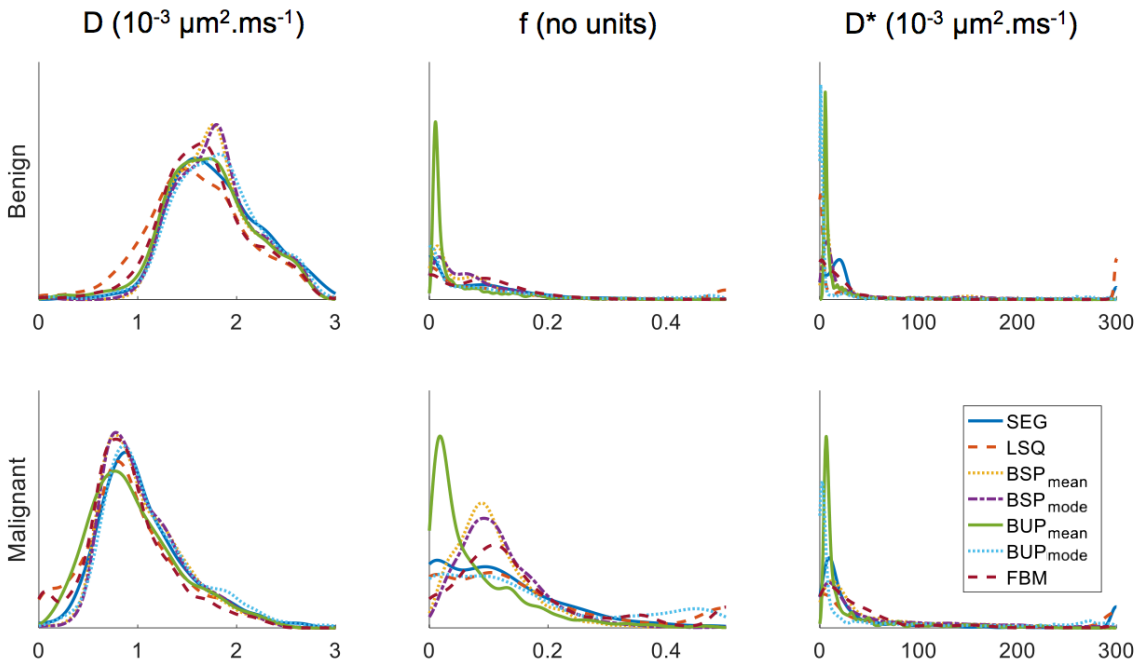

Figure 2. Histograms showing distributions of the estimated IVIM parameters obtained using different fitting methods, for all voxels in benign and malignant lesion sub-cohorts. Some statistical differences appear between distributions of different fitting methods, but the dominant general trend of increased *f* and decreased *D* for malignant lesions, which forms the basis of breast lesion classification using IVIM, is preserved.

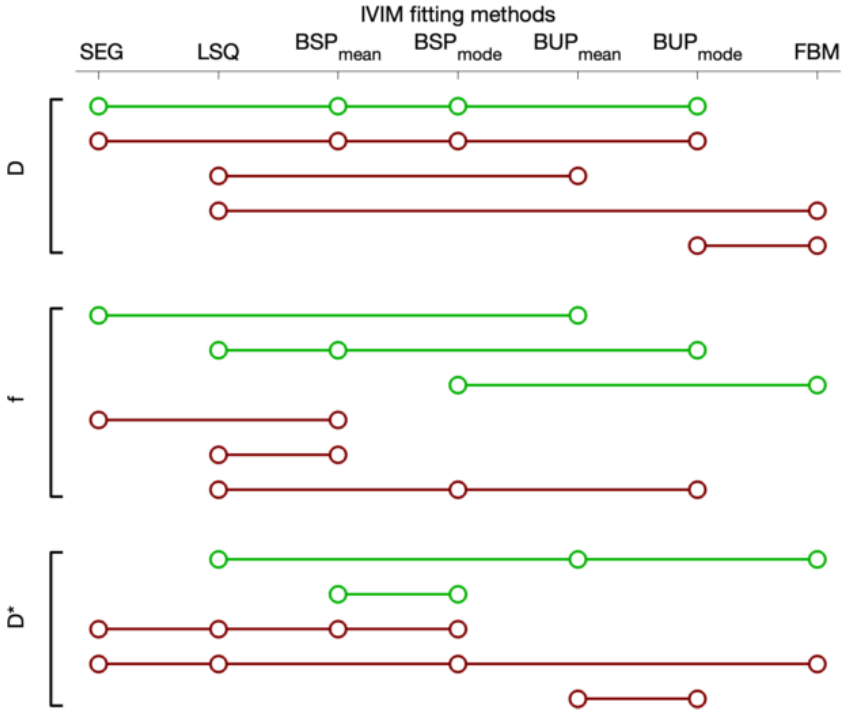

Figure 3. Connections indicate non-significant p-values (Mann-Whitney test with Bonferroni correction; $p > 0.05$) for comparison of IVIM parameter values obtained by different fitting methods; presented for benign (green) and malignant (red) subgroups. These data illustrate that the influence of the fitting methods is often statistically detectable, although minor in comparison with values compared across malignant/benign groups (Figure 2).

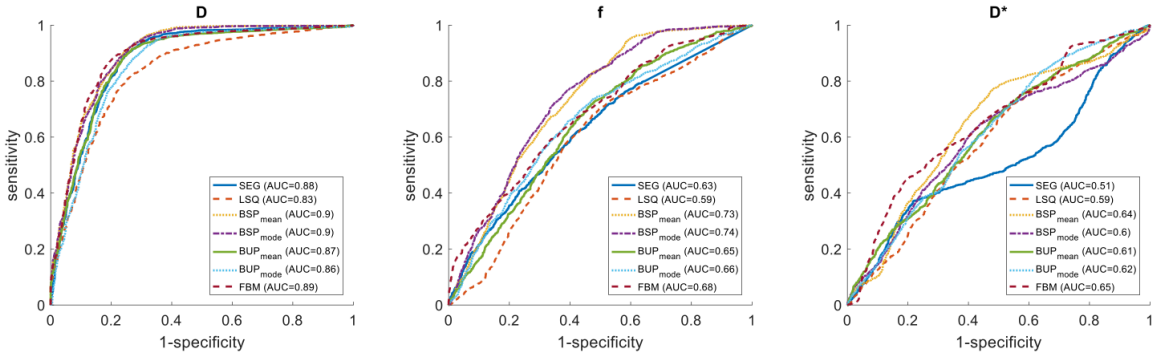

Figure 4a ROC curves for differentiation of malignant/benign tumors using individual IVIM parameters from different fitting methods. In line with expectation, *D* outperforms *f* and *D** in all methods, and *D* achieves a minimum AUC of 0.83 with LSQ method (interestingly, SEG outperforms LSQ for both *D* and *f*). Bayesian methods, specifically BSP and FBM, achieve the highest AUC of 0.9 and 0.89 respectively.

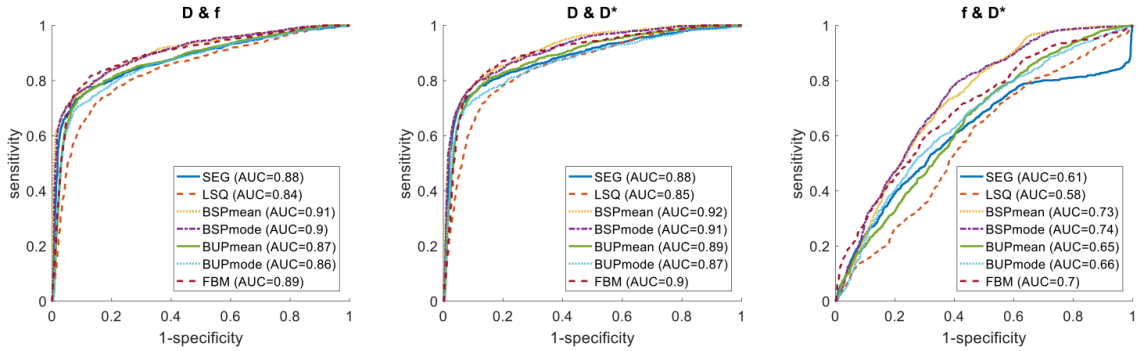

Figure 4b ROC curves, with corresponding AUC values, for logistic regression combination pairs of IVIM parameters in classifying malignant vs. benign breast tumors. While *D* is the single best parameter, the addition of *f* or *D** shows increased discriminatory power.

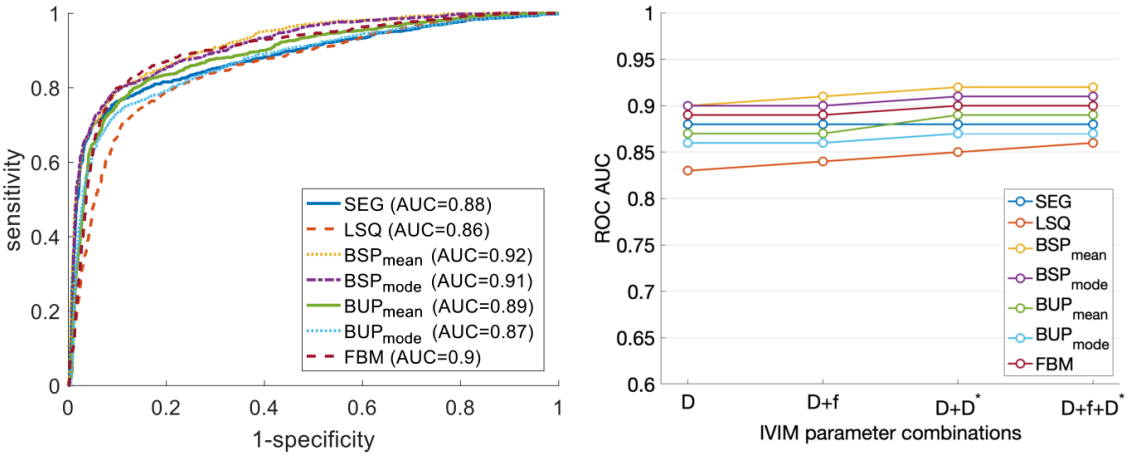

Figure 4c Corresponding ROC curves where logistic regression allows for combining all parameters (left); adding contributions from *f* and *D** to *D* increases AUC by approximately 0.017 (for all methods except SEG). Graphical representation of ROC AUCs for each method and combination are shown (right).

Table 1: Patient characteristics for the study cohort.

Lesion Classification	Malignant	Benign
Number	28	23
Mean patient age (years)	53.7 (29 – 75)	29.9 (21 – 53)
Mean tumor volume (cm^3)	1.5 (0.2 – 4.25)	8.1 (0.1 – 104.1)
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)	Fibroadenoma (14) Phyllodes (2) Fibroadenomatosis (1) Adenosis (1) No histologic analysis available (5)
Histologic grade	1:6, 2:9, 3:9, 2/3:2, not analyzed: 2	

Table 2. Correlation of parameters between methods and across all lesions. In general, D is highly correlated across all methods, whereas far less correlation is seen for f and D*.

D	SEG	LSQ	BSP_{mean}	BSP_{mode}	BUP_{mean}	BUP_{mode}	FBM
SEG	1	0.84	0.9	0.89	0.91	0.89	0.85
LSQ		1	0.88	0.89	0.92	0.88	0.88
BSP_{mean}			1	0.99	0.96	0.96	0.91
BSP_{mode}				1	0.96	0.96	0.91
BUP_{mean}					1	0.97	0.88
BUP_{mode}						1	0.92
FBM							1

f	SEG	LSQ	BSP_{mean}	BSP_{mode}	BUP_{mean}	BUP_{mode}	FBM
SEG	1	0.46	0.62	0.63	0.68	0.55	0.55
LSQ		1	0.35	0.36	0.41	0.62	0.32
BSP_{mean}			1	0.94	0.75	0.5	0.55
BSP_{mode}				1	0.72	0.53	0.6
BUP_{mean}					1	0.59	0.45
BUP_{mode}						1	0.46
FBM							1

D*	SEG	LSQ	BSP_{mean}	BSP_{mode}	BUP_{mean}	BUP_{mode}	FBM
SEG	1	0.48	0.36	0.43	0.4	0.57	0.3
LSQ		1	0.42	0.49	0.44	0.7	0.33
BSP_{mean}			1	0.92	0.68	0.59	0.54
BSP_{mode}				1	0.7	0.68	0.49
BUP_{mean}					1	0.68	0.38
BUP_{mode}						1	0.32
FBM							1

Table 3. Average coefficients of variation (CV) for IVIM parameters across lesions (thus also including true variation). For every approach, CV is lowest for *D* and highest for *D**. The pseudodiffusion parameters exhibit lower CVs in malign lesions than in benign lesions, which may be a result of a more detectable/defined contribution to the IVIM signal. The pseudodiffusion parameter estimates from BSP and FBM possess lower CVs than those obtained from the conventional methods.

Malignant	D	f	D*
SEG	0.28	0.88	1.49
LSQ	0.39	0.98	1.31
BSPmean	0.21	0.53	0.88
BSPmode	0.21	0.49	1.09
BUPmean	0.32	1	1.14
BUPmode	0.27	0.97	1.58
FBM	0.35	0.62	0.84

Benign	D	f	D*
SEG	0.17	1.09	1.57
LSQ	0.26	1.2	1.45
BSPmean	0.11	0.62	0.77
BSPmode	0.11	0.67	0.98
BUPmean	0.18	1.07	1.26
BUPmode	0.17	1.27	1.79
FBM	0.21	0.75	1.01