FULL PAPER

Modeling the diffusion-weighted imaging signal for breast lesions in the b = 200 to 3000 s/mm^2 range: quality of fit and classification accuracy for different representations

Igor Vidić[1] | Liv Egnell[1,2] | Neil P. Jerome[2,3] | Nathan S. White[4,5,6] | Roshan Karunamuni[7] | Rebecca Rakow-Penner[4] | Anders M. Dale[4,8] | Tone F. Bathen[2,3] | Pål Erik Goa[1,2]

[1]Department of Physics, NTNU–Norwegian University of Science and Technology, Trondheim, Norway

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

[3]Department of Circulation and Medical Imaging, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

[4]Department of Radiology, University of California San Diego, La Jolla, California

[5]Center for Multimodal Imaging and Genetics, University of California, San Diego, La Jolla, California

[6]HealthLytix Inc., San Diego, California

[7]Department of Radiation Medicine and Applied Sciences, University of California, San Diego, La Jolla, California

[8]Department of Neurosciences, University of California, San Diego, La Jolla, California

Correspondence
Pål Erik Goa, Department of Physics, NTNU–Norwegian University of Science and Technology, Høgskoleringen 5, Realfagbygget D5-170, NO-7491 Trondheim, Norway.
Email: pal.e.goa@ntnu.no

Funding information
Mid-Norway Regional Health Authority, Grant/Award Number: 46084400

Purpose: To evaluate different non-Gaussian representations for the diffusion-weighted imaging (DWI) signal in the b-value range 200 to 3000 s/mm^2 in benign and malignant breast lesions.

Methods: Forty-three patients diagnosed with benign (n = 18) or malignant (n = 25) tumors of the breast underwent DWI (b-values 200, 600, 1200, 1800, 2400, and 3000 s/mm^2). Six different representations were fit to the average signal from regions of interest (ROIs) at different b-value ranges. Quality of fit was assessed by the corrected Akaike information criterion (AICc), and the Friedman test was used for assessing representation ranks. The area under the curve (AUC) of receiver operating characteristic curves were used to evaluate the power of derived parameters to differentiate between malignant and benign lesions. The lesion ROI was divided in central and peripheral parts to assess potential effect of heterogeneity. Sensitivity to noise-floor correction was also evaluated.

Results: The Padé exponent was ranked as the best based on AICc, whereas 3 models (kurtosis, fractional, and biexponential) achieved the highest AUC = 0.99 for lesion differentiation. The monoexponential model at b$_{max}$ = 600 s/mm^2 already provides AUC = 0.96, with considerably shorter acquisition time and simpler analysis. Significant differences between central and peripheral parts of lesions were found in malignant lesions. The mono- and biexponential models were most stable against varying degrees of noise-floor correction.

Conclusion: Non-Gaussian representations are required for fitting of the DWI curve at high b-values in breast lesions. However, the added clinical value from the high b-value data for differentiation of benign and malignant lesions is not clear.

1 | INTRODUCTION

Diffusion-weighted magnetic resonance imaging (DWI)[1] is sensitive to the random motion of water molecules. In the case of Gaussian, or free diffusion, the DWI signal can be modeled as a monoexponential decay, as a function of the degree of applied diffusion weighting (b-value) and the underlying diffusion coefficient. In biological tissue, the self-diffusion of water is known to be more complicated, but the monoexponential model is still commonly applied to fit data arising from the application of ≥ 2 low and medium b-values, in the range 0 to 1000 s/mm^2, providing an apparent diffusion coefficient (ADC). Given that the true b-value dependence is not monoexponential, the measured ADC will not only depend on the tissue diffusion properties, but it is also influenced by the chosen number and magnitude of the b-values in the acquisition,[2] as well as other acquisition parameters such as the TE and diffusion time (Δ).[3]

Given the complexity of DWI signal decay in vivo, it is useful to separate discussion of diffusion in terms of the b-value range involved. At low b-values, typically in the range 0 to 200 s/mm^2, a faster decay of DWI signal is commonly observed compared to that at higher b-values (b > 200 s/mm^2). The intravoxel incoherent motion (IVIM) model attempts to capture this effect through an additional exponential b-value dependence[4] that is attributed to incoherent blood flow (perfusion) in a randomly oriented capillary network, manifesting as a faster, pseudo-diffusion process. In the b-value range of 200 to 1000 s/mm^2, b-value dependence is difficult to distinguish from monoexponential, and for the typical signal-to-noise ratios (SNRs) available in clinical DWI protocols, significant deviation from this model is usually not detected. When this b-value range is extended, for example up to 2000 to 3000 s/mm^2, it is well documented that a monoexponential function ceases to be a good approximator for the observed data.[5] This effect has attracted considerable attention over the years, particularly in the brain,[6-11] where it was shown that a biexponential representation, without the pseudo-diffusion interpretation of the IVIM model, in the higher b-value regime provided good fits to the diffusion data.[12-14] Early efforts to link this 2-compartment representation to biophysical compartments in the brain were not successful,[12,14,15] and it was later shown that other mathematical representations,

without clear biophysical interpretation, could fit the data equally well or even better using fewer parameters.[7,8]

Many mathematical functions have been proposed to describe the DWI signal decay: the diffusion kurtosis (DKI) representation,[9] Padé exponent (PE) representation,[8] stretched exponential model,[11] statistical model,[16] fractional order calculus model,[17] and biexponential model.[18] Some functions attempt to describe the signal as a function of variables related to physical properties of the system, and as such are true models (e.g., the biexponential model). Others are based on mathematical approximations and may be useful in a clinical context, but lack a biophysical foundation (e.g., DKI). Such functions are merely signal interpolations and are inherently unsuitable for extrapolation of signal curves, given that their parameters are not verifiable by other techniques. The importance of distinguishing between these two concepts is discussed by Novikov et al[19] and is highly relevant for model validation; accordingly, we use the word "representation" for both of these categories and reserve the word "model" exclusively for the first category of functions, which represent actual models of a physical system.

There is currently strong interest in using DWI to better characterize the microstructure of breast cancer tissue,[20-22] with a commensurate need to study and evaluate candidate biophysical models. One necessary, though not sufficient, condition for model validation is to assess its quality of fit compared to other representations.[23] In addition, aside from the question of biophysical interpretation, different diffusion representations should be considered on an organ-specific basis, particularly in relation to robustness and clinical value of the extracted parameters.

The aim of the current study was to evaluate the quality of fit for a chosen set of mathematical representations (with or without a corresponding biophysical model) and to determine the representation that best fits DWI of the breast in an available b-value range up to 3000 s/mm^2 in a high-SNR situation using region of interest (ROI) mean signal intensities. Furthermore, the power of the obtained diffusion parameters to differentiate between benign and malignant breast lesions was evaluated. We also separated the ROI into a peripheral and central region to see whether heterogeneity related to proximity to tumor edge influenced the models' parameter estimates and their ability to fit the signal decay. The effect of

varying the degree of noise-floor correction was examined for the different signal representations. Finally, all models were fitted to individual single voxels to evaluate the classification accuracy for this approach compared to the ROI-mean case.

2 | METHODS

2.1 | 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 before enrollment. The recruitment of patients for this study began in August 2014 and ended in August 2016. Patients with recently diagnosed malignant or benign tumors in our institution were able to volunteer for an additional MR examination. Only patients potentially eligible for primary surgical treatment were recruited, and no patients receiving neoadjuvant chemotherapy were included. Exclusion criteria were general contraindications for MRI (pacemaker, aneurysm clips, any form of metal in the body, or severe claustrophobia) and impaired renal function defined as estimated glomerular filtration rate <30 mL/min/1.73 m^2.

Following MR examination, patients with malignant tumors underwent surgery and histopathological analysis was performed on the resected mass. Categorization of benign tumors was done by histopathological analysis on core needle biopsies or resected tissue, if the tumor was surgically removed. For benign lesions where a biopsy was not requested by the radiologist, diagnosis was based on patient history, which included either radiographical mammography, ultrasonography, or a previous clinical MR examination with at least 6 months' follow-up from the time of recruitment. Forty-seven patients were recruited. The cohort included in this study is a subgroup of the cohort previously reported on in a study by Vidic et al, where a machine-learning approach was used to analyze parameter histograms derived from a separate IVIM DWI protocol.[24]

2.2 | MRI protocol

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

Fat-suppressed (n = 24 using "FatSat" and n = 23 using spectral attenuated inversion recovery [SPAIR]) unilateral sagittal DWI was acquired using a Stejskal–Tanner spin-echo echo-planar imaging sequence with: TR 10,600 (n = 24) or 11,800 ms (n = 23), TE 88 ms, diffusion gradient duration δ = 31.1 ms, diffusion time Δ = 40.7 ms, 90 × 90 matrix,

2 × 2 mm in-plane resolution, slice thickness 2.5 mm, 60 slices, generalized autocalibrating partially parallel acquisition factor 2, number of excitations = 1, b-values: 0, 200, 600, 1200, 1800, 2400, and 3000 s/mm^2 in 6 directions, total scan time 7 minutes 46 seconds (FatSat) and 8 minutes 39 seconds (SPAIR). The protocol included 1 additional, geometry-matched, non-diffusion-weighted (b = 0 s/mm^2) series with reversed phase-encoding direction for implementation of distortion correction arising from susceptibility boundaries.[25]

Patients also underwent dynamic contrast-enhanced (DCE) MRI. DCE scans consisted of 3D, T$_1$-weighted, non-fat-suppressed, gradient echo sequence (TR/TE 5.82/2.18 ms, flip angle 15°, 256 × 256 matrix, in-plane resolution 0.7 × 0.7 mm, slice thickness 2.5 mm) acquisitions, collected pre-contrast, and at 7 consecutive time points (with temporal resolution of 1 minute) after administration of contrast agent. DCE MR images were used for guidance of ROI selection in the DWI data.

2.3 | Data analysis

DWI images were visually inspected for the presence of artefacts. All DWI data were corrected for eddy current artefacts, motion,[6] geometrical distortion,[25,26] and noise-floor. Noise-floor correction was done by applying a similar scheme as in a previous work.[27] A corrected and normalized signal intensity S$_c$ was calculated by Equation 1:

$$S_c = \sqrt{S_m^2 - NF^2}/NF \tag{1}$$

where *NF* is the noise-floor correction factor and given by $NF^2 = k \cdot BK^2$, where *k* is a scaling factor and *BK* is the background intensity defined as the mean intensity inside a ROI covering the whole breast at b = 3000 s/mm^2 in each patient. Normalization by dividing with *NF* was done in order to remove the effect of different image intensity scaling between FatSat and SPAIR protocols. S$_c$ was set to 0 if $S_m < NF$. The effect of varying the noise-floor correction scaling factor was included in the first step of the analysis by evaluating the quality of fit for diffusion representations that included the lowest signal-to-noise-floor images (b = 3000 s/mm^2) at increasing value of *k*, and the resulting empirical value of *k* = 1 was used for all subsequent analysis.

Where multiple lesions were present in the same breast, the largest was selected for analysis. Lesions were outlined by 3D ROIs in the b = 600 s/mm^2 trace image, guided by reference to a postcontrast DCE image to locate the lesion, by a scientist (I.V.) with 3 years' research experience in breast cancer MRI, supervised and validated by an experienced radiologist (20 years' experience in breast radiology). In the

presence of non-mass-like lesions, only the mass part was included in ROI. Mean ROI signal for each b-value and diffusion direction was calculated and used for the main analysis, which included both quality of fit and classification accuracy. In addition, classification accuracy was also assessed for pooled single voxels.

ROIs for the lesion center were automatically computed from the full ROI by applying a 2D erosion filter on the original ROI using a disk with a radius of 1 voxel in each slice. Complementary ROIs for the lesion periphery were created by subtracting the central ROIs from the corresponding full ROIs.

The single voxel signal standard deviation (STD) was calculated from a difference image between two b = 0 s/mm^2 acquisitions for each patient, where the STD was calculated across the ROI voxels in the difference image and divided by $\sqrt{2}$. Mean noise-floor-to-STD ratio across all patients was then calculated.

The additional single voxel analysis was performed by first collecting all voxels from the benign and malignant lesions into 2 separate matrices. The 18 benign lesions counted in total 17,293 voxels, whereas the 25 malignant lesions counted 6278 voxels. After fitting each diffusion representation to each voxel individually, receiver operating characteristic (ROC) analysis was performed.

2.4 | Diffusion representations

To avoid possible influence from perfusion, the b = 0 s/mm^2 ROI signal was excluded from the fitting, and all b-values from b = 200 s/mm^2 up to a maximum b-value (b$_{max}$) were included in the fit. Three different values of b$_{max}$ were used: b$_{max}$ = 600 s/mm^2, similar to a standard clinical DWI protocol, b$_{max}$ = 1800 s/mm^2, representing a medium-range non-Gaussian protocol, and b$_{max}$ = 3000 s/mm^2, representing a wide-range non-Gaussian protocol. Different diffusion representations were applied in these 3 different b-value ranges as specified in more detail below. All representations were fitted to the DWI signal using Matlab's (2016b; The MathWorks, Inc., Natick, MA) in-built nonlinear least squares solver with Trust region optimizer.

The following diffusion signal representations were used:

Monoexponential representation calculated by fitting the equation (Equation 2)[1]:

$$S_b = S_0 \cdot e^{-b \cdot ADC} \tag{2}$$

where ADC is the apparent diffusion coefficient. The monoexponential model was only used at the lower end of the b-value range, with b$_{max}$ = 600 s/mm^2.

DKI representation (Equation 3)[9]:

$$S_b = S_0 \cdot e^{-b \cdot D_k + \frac{1}{6}b^2 \cdot D_k^2 \cdot K} \tag{3}$$

where D_k measures diffusivity while K, the apparent kurtosis excess, is a measure of peakiness of the diffusion displacement probability distribution. DKI is derived from the cumulant expansion, truncated after the quadratic term, and is only valid within the convergence radius of the function,[7,8] beyond which DKI does not perform well. As a consequence, the DKI model was only applied in the medium b-value range with b$_{max}$ = 1800 s/mm^2.

PE representation (Equation 4)[8]:

$$S_b = S_0 \cdot e^{\frac{-D_P \cdot b}{1 + \delta_P \cdot b}} \tag{4}$$

where δ_P quantifies non-Gaussianity and D_P diffusivity. Note, that when $\delta_P = 0$, the decay is monoexponential. Given that the PE is a polynomial approximation, the PE representation does not suffer from the convergence issues as DKI.[8] The PE representation was applied with b$_{max}$ = 1800 and 3000 s/mm^2.

It is important to note that both the DKI and Padé approximation are mathematical series expansions that can be used for approximating any analytical function. This also means that both are purely mathematical representations that do not include biophysical modeling or require any assumptions about tissue microstructure.

Stretched exponential representation (Equation 5)[8]:

$$S_b = S_0 \cdot e^{-(b \cdot DDC)^\alpha} \tag{5}$$

where DDC represents the distributed diffusion coefficient and α the heterogeneity index, which varies between 0 and 1. The limiting value $\alpha = 1$ corresponds to monoexponential diffusion. This representation attempts to capture anomalous diffusion[28] and assumes that the medium is scale invariant (fractal). The stretched exponential representation was applied with b$_{max}$ = 1800 and 3000 s/mm^2.

Statistical representation (Equation 6)[16]:

$$S_b = S_0 \cdot \frac{1 + \Phi\left(\frac{ADC_{peak}}{\sigma\sqrt{2}} - \frac{b\sigma}{\sqrt{2}}\right)}{1 + \Phi\left(\frac{ADC_{peak}}{\sigma\sqrt{2}}\right)} e^{-b \cdot ADC_{peak} + \frac{1}{2}b^2\sigma^2} \tag{6}$$

which assumes a Gaussian distribution of diffusion coefficients where ADC_{peak} represents the peak of the distribution, σ is the width of the distribution, and Φ is the error function. b$_{max}$ = 1800 and 3000 s/mm^2.

Fractional calculus representation in the case of Stejskal–Tanner gradients (Equation 7)[17]:

$$S_b = S_0 \cdot e^{-(b^* \cdot D^*)^\beta} \tag{7}$$

which has the same mathematical form as the stretched exponential representation, but with the following definitions (Equations 8 and 9):

$$b^* = \left(\gamma G_d \delta\right)^2 \left(\Delta - \frac{2\beta - 1}{2\beta + 1}\delta\right) \tag{8}$$

$$D^* = D^{\frac{1}{\beta}} \mu^{2\left(1-\frac{1}{\beta}\right)} \left(\Delta - \frac{2\beta - 1}{2\beta + 1}\delta\right)^{\frac{1}{\beta}-1} \tag{9}$$

where G_d is the diffusion gradient amplitude, δ and Δ are the diffusion gradient pulse lobe duration and separation, and D is the diffusion coefficient. β is the fractional order derivative in space, and μ is a spatial parameter (in units of μm). $b_{max} = 1800$ and 3000 s/mm^2.

Biexponential representation (Equation 10)[18]:

$$S_b = S_0 \left(f \cdot e^{-b \cdot D_1} + (1-f) \cdot e^{-b \cdot D_2}\right) \tag{10}$$

where, if interpreted as a 2-component model, the signal fraction f corresponds to the fraction of "fast" diffusion component with the apparent diffusion coefficient D_1, while $(1 - f)$ and D_2 represents the fraction and apparent diffusion coefficient of the "slow" component[29] $(D_1 > D_2)$. The biexponential model presented here is mathematically identical to the IVIM model. However, a different b-value range is involved and therefore also a different biophysical model. $b_{max} = 1800$ and 3000 s/mm^2.

2.5 | Statistical analysis

The corrected Akaike information criterion (AICc) was used to assess the curve fitting performance of the selected diffusion representations. AICc is a statistical information criterion to assess the quality of fit of a representation to the observed data (Equation 11):

$$AICc = \frac{2 \cdot k \cdot n}{n - k - 1} + n \cdot \ln\frac{RSS}{n} \tag{11}$$

where n is the number of data points, RSS the sum-of-squares residuals, and k the number of variables. AICc represents a trade-off between the goodness of fit and complexity of the representation. For a given set of candidate representations for the same data points, the preferred representation is the one with the minimum AICc value. Only the relative value of AICc is meaningful, and for this reason all results are reported in terms of $\Delta AICc = AICc - AICc_{min}$, where $AICc_{min}$ is the AICc value of the representation with the lowest AICc in the given comparison. Compared to the standard AIC, AICc includes a correction for small sample sizes. Calculation of AICc and all the image analysis and statistical testing was done in Matlab (2016b; The MathWorks).

The Friedman test was used to test whether the ranks of the AICc values from different representations differed significantly. The AUC of the ROC curves were used to assess the power of extracted parameters to differentiate between malignant and benign tumors. A signed-rank test was used to compare parameter values for data within central ROI versus peripheral ROI.

3 | RESULTS

From the 47 patients recruited, DWI data were successfully collected in 43 patients. Four data sets were excluded from analysis because of Nyquist ghosting artefacts. Among the 43 lesions included in the analysis, 18 lesions were benign and 25 malignant (clinical data are reported in Table 1). One example breast diffusion –weighted image with central and peripheral ROIs is illustrated in Figure 1.

TABLE 1 Patient cohort

Characteristics	Result
Malignant	
No. of lesions	25
Mean patient age, years (range)	53.3 (29–75)
Mean tumor volume, cm^3 (range)	2.5 (0.5–5.8)
Histological type	
Invasive ductal carcinoma	9
Invasive ductal carcinoma with ductal carcinoma in situ	10
Medullary carcinoma with ductal carcinoma in situ	2
Invasive lobular carcinoma	2
Mucinous carcinoma with ductal carcinoma in situ	1
Papillary carcinoma	1
Histological grade	
1	5
2	7
3	8
2/3	1
Not analyzed	4
Benign	
No. of lesions	18
Mean patient age, years (range)	27.2 (20–48)
Mean tumor volume, cm^3 (range)	9.6 (0.5–88.4)
Histological type	
Fibroadenoma	12
Phyllodes	2
Fibroadenomatosis	1
Adenosis	1
No histological analysis available	2

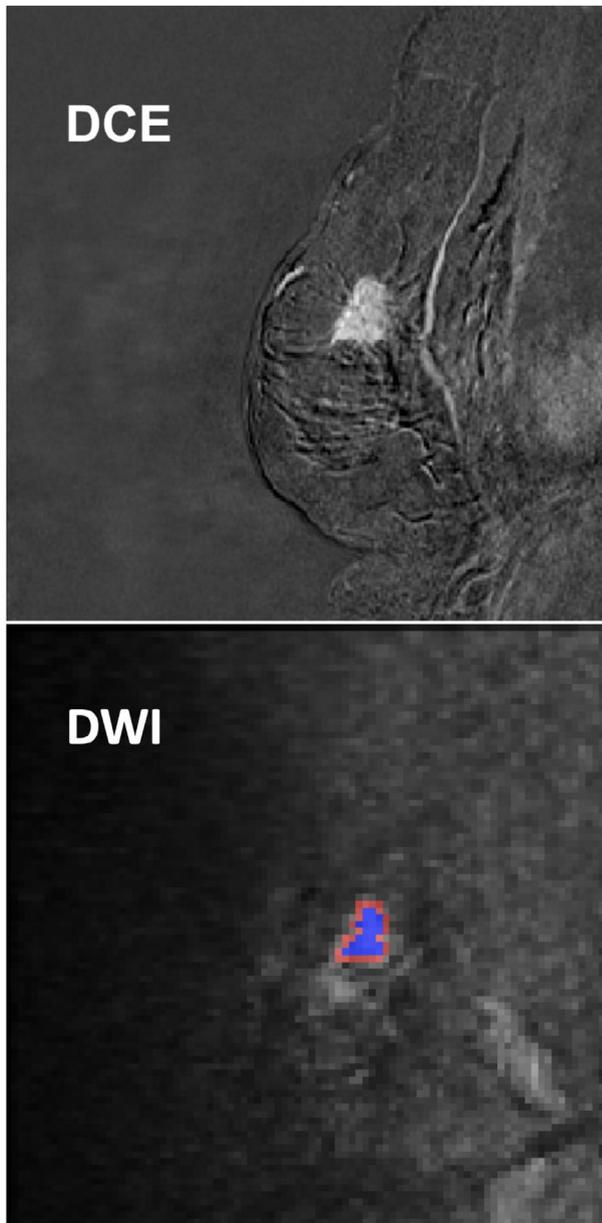**FIGURE 1** Example of DCE and b = 600s/mm² DWI image with ROI, showing peripheral (red) and central parts (blue) of ROI constituting complete ROI (patient aged 62 years diagnosed with invasive ductal carcinoma)
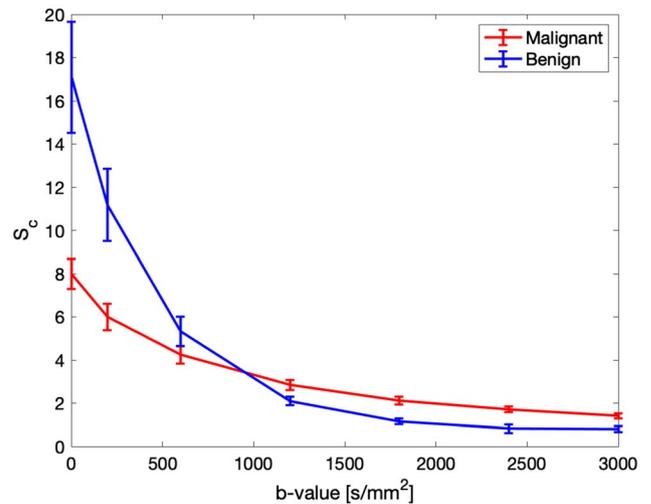

FIGURE 2 Median corrected signal-to-noise-floor (S_c) for 1 example benign and malignant lesion as a function of b-value. Error bars indicate $IQR/\sqrt{n}$. Note that to avoid any perfusion influence, the b = 0 s/mm² signal was not included in the fit

Signal decays for 1 benign and 1 malignant lesion are shown in Figure 2.

The results from varying the noise-floor correction factor *NF* are shown in Figure 3 for b_{max} = 3000 s/mm². A systematic variation in the quality of fit as measured by AICc can be seen; for the biexponential model, the AICc is monotonically increasing with the scaling factor *k*, indicating a reduced quality of fit with increasing *NF*. For several models, a minimum AICc is found for *k* = 1. Already at *k* = 0.5, a few measurement points are found to be below NF, and at *k* = 1 this amounts to almost 5% of the measurement points.

The median noise-floor-to-STD was 1.41. This means that the single voxel SNR at b = 0 s/mm² for this DWI protocol was around 20 for a typical benign lesion and 10 for a malignant lesion, based on the noise-floor–normalized S_0 values fitted for the monoexponential model reported in Table 2 (14.28 and 7.15, respectively).

Median values and interquartile ranges (IQRs) in benign and malignant lesions for all fitted parameters are reported in Table 2. Friedman test mean rankings of the different representations regarding the quality of fit as measured with AICc are reported in Table 3. The Padé representation had the best overall mean ranks of 2.02 for b_{max} = 3000 s/mm² and 1.93 for or b_{max} = 1800 s/mm². Mean ΔAICc for the different models and b_{max}-values are shown in Table 4, with lower ΔAICc scores indicating a better quality of fit.

Plots of the average residuals as a function of b-value (Figure 4) illustrate that the PE representation best fitted the mean ROI signal as a function of b-value, but also systematic residuals from other representations.

AUC for all model parameters to differentiate between malignant and benign lesions at different values of b_{max} are shown in Table 5. All representations reached AUC > 0.9 for at least 1 parameter. In 3 cases, AUC = 0.99 was reached; for the *K* parameter in the kurtosis representation at b_{max} = 1800 s/mm², and for *D* in the fractional calculus model and *f* in the biexponential model at b_{max} = 3000 s/mm².

There were significant differences between central and peripheral ROIs in terms of S_0 in both malignant and benign lesions while only in malignant lesions for the diffusion related parameters (see Supporting Information Tables S1 and S2).

The results from the additional single voxel analysis are summarized in Supporting Information Table S4. Highest
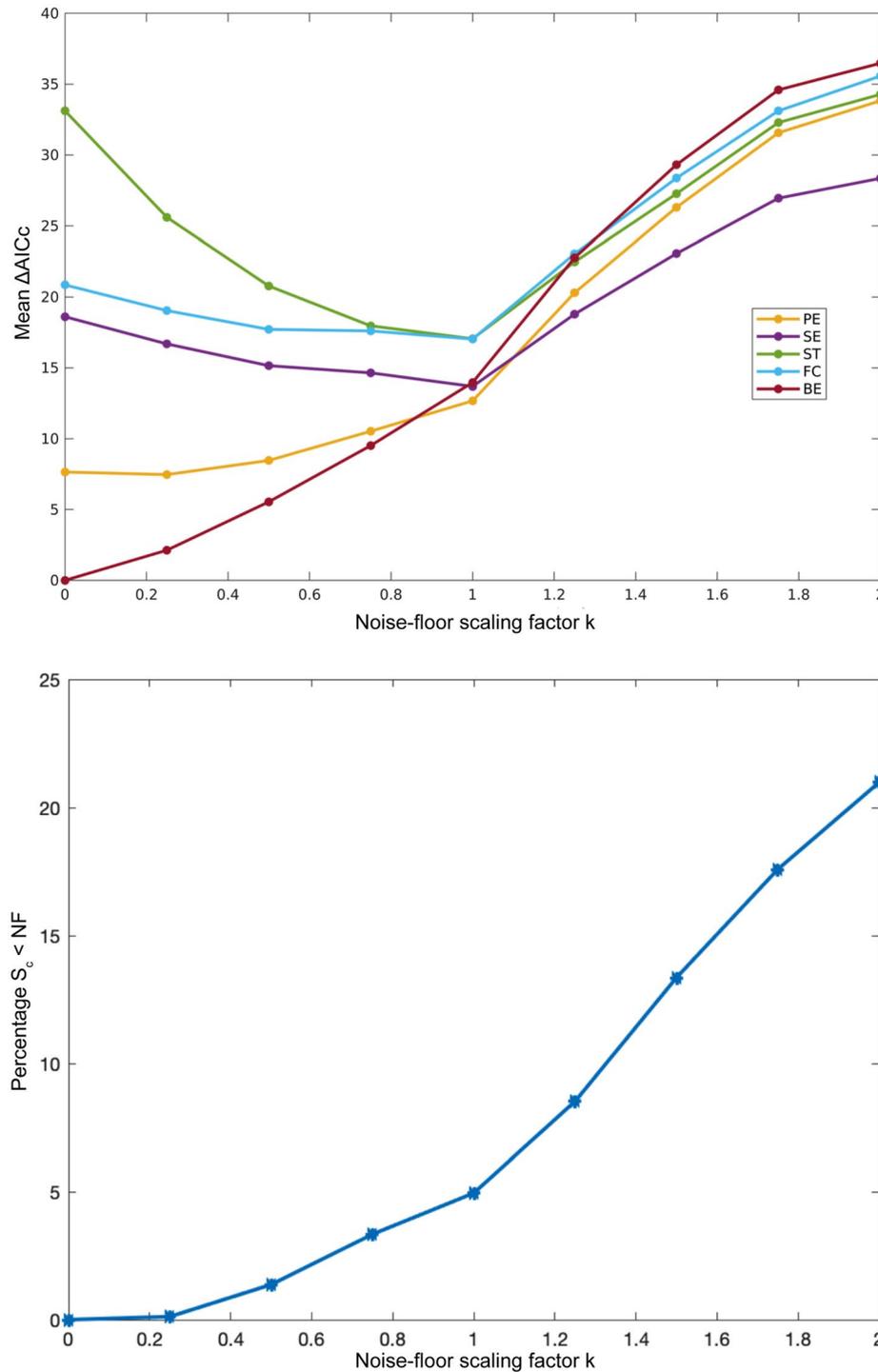
FIGURE 3 (A) Effect of different choices of noise-floor correction factor (*NF*) on the mean AICc for the models applied at the full range of b-values. (B) Percentage number of measurements with intensity lower than *NF*. $NF^2 = k \cdot BK^2$, where *BK* is the mean intensity inside the breast at b = 3000 s/mm². For the nominal scaling factor of $k = 1$, the different models have similar AICc, and approximately 95% of voxels are included. BE, biexponential; FC, fractional calculus; PE, Padé exponent; SE, stretched exponential; ST, statistical

classification accuracy was observed for the fractional model, with an AUC of 0.95. Also, the statistical model provided a higher AUC (0.94) than the monoexponential model (0.92). The other non-Gaussian representations provided lower AUC than the monoexponential model.

4 | DISCUSSION

The motivation behind this study was twofold; to investigate how different non-Gaussian diffusion representations fit the b-value dependence of the DWI signal in breast lesions

TABLE 2 Parameter values, median with IQR for different values of b_{max}

		$b_{max} = 600$ s/mm^2							
		Malignant		Benign					
		Median	IQR	Median	IQR				
Monoexp.	S_0 [S/NF]	7,15	3,54	14,28	10,66				
	ADC [µm^2/ms]	0,97	0,23	1,82	0,28				
		$b_{max} = 1800$ s/mm^2				$b_{max} = 3000$ s/mm^2			
		Malignant		Benign		Malignant		Benign	
		Median	IQR	Median	IQR	Median	IQR	Median	IQR
Kurtosis	S_0 [S/NF]	7,36	3,72	14,82	11,47				
	K	1,01	0,29	0,39	0,2				
	D_k [µm^2/ms]	1,14	0,27	1,92	0,38				
Padé	S_0 [S/NF]	7,61	3,88	15,03	11,48	7,76	3,76	15,34	11,39
	δ_P	0,41	0,15	0,19	0,13	0,37	0,11	0,21	0,23
	D_p [µm^2/ms]	1,36	0,35	2,06	0,42	1,29	0,30	2,11	0,53
Stretched exp.	S_0 [S/NF]	10,18	5,52	17,99	14,62	10,96	5,40	23,54	17,75
	α	0,57	0,11	0,78	0,13	0,51	0,10	0,72	0,23
	DDC [µm^2/ms]	1,36	0,70	2,23	0,90	1,57	0,74	2,51	1,02
Statistical	S_0 [S/NF]	7,30	3,44	14,89	11,31	7,17	3,30	14,71	11,21
	σ	1,00	0,00	0,77	0,28	1,00	0,01	0,77	0,57
	ADC_{peak} [µm^2/ms]	0,70	0,56	1,93	0,42	0,69	0,62	1,93	0,39
Fractional	S_0 [S/NF]	10,18	5,52	17,99	14,62	10,96	5,40	23,54	17,75
	μ [µm]	4,84	0,15	5	0,46	4,66	0,18	4,84	0,39
	β	0,57	0,11	0,78	0,13	0,51	0,10	0,72	0,23
	D [µm^2/ms]	0,8	0,2	1,62	0,23	0,78	0,2	1,57	0,28
Biexp.	S_0 [S/NF]	7,51	3,65	14,52	10,95	7,38	3,70	14,39	10,91
	f	0,61	0,23	0,92	0,09	0,56	0,16	0,95	0,09
	D_1 [µm^2/ms]	2,01	0,67	2,02	0,33	1,89	0,40	1,90	0,28
	D_2 [µm^2/ms]	0,38	0,27	0,49	0,15	0,33	0,22	0,49	0,43

and to investigate the possible added clinical value of a high b-value protocol.

Regarding the first motivation, the results are interesting and, in some cases, also somewhat surprising. In the noise-floor–corrected case, the kurtosis and Padé representations, both purely mathematically motivated representations, are ranked with the best quality of fit determined by AICc for $b_{max} = 1800$ s/mm^2 (see Table 3). For $b_{max} = 3000$ s/mm^2, the Padé and stretched exponential representations are ranked as best. The representations with 4 free parameters, the biexponential and the fractional calculus, did not perform well in terms of AICcs. The fractional calculus representation had the lowest quality of fit for both values of b_{max}, whereas the biexponential was ranked as third for the higher b_{max}. Higher AICc values for these representations were expected as a consequence of the introduced penalty for the number of parameters in the AICc calculation.

The fractional calculus representation, while having more parameters, also had larger residuals than the Padé and stretched exponential representations (Figure 4); so the higher AICc was unsurprising also from that point of view. Conversely, residuals for the biexponential model were the second lowest (after Padé; Figure 4).

Given the relatively low signal-to-noise-floor at the highest b-values, it is important to handle the noise-floor correctly, or at least to know how the results are influenced by incorrect noise-floor correction. Commonly, the noise-floor (NF) value is estimated based on image SNR from repeated measurements. However, for multielement coils combined with parallel imaging methods, the relationship between NF and SNR is nontrivial. In this article, we chose to estimate NF directly from the b = 3000 s/mm^2, assuming effectively no real signal from the breast at that b-value and that the small number of lesion voxels that retain

TABLE 3 Mean ranks, regarding the quality of fit as measured with AICc, obtained from Friedman test for benign and malignant cases in complete, central, and peripheral part of ROI

	DKI	PE	SE	ST	FC	BE
All (b_{max} = 1800 s/mm^2)	2.26	1.93	2.74	3.70	5.63	4.74
All (b_{max} = 3000 s/mm^2)		2.02	2.04	3.77	3.95	3.26
Benign (b_{max} = 1800 s/mm^2)	2.37	1.88	2.72	3.55	5.51	4.91
Benign (b_{max} = 3000 s/mm^2)		2.07	2.16	3.34	4.11	3.30
Malignant (b_{max} = 1800 s/mm^2)	2.05	2.14	2.79	3.60	5.60	4.81
Malignant (b_{max} = 3000 s/mm^2)		1.91	1.98	3.93	3.79	3.40

Mean ranks across all models differ with high significance ($P < .001$).

Abbreviations: BE, biexponential; FC, fractional calculus; ST, statistical.

TABLE 4 Mean ΔAICc for the different models at the 2 different choices of b_{max}

	b_{max} [s/mm^2]	
ΔAICc	1800	3000
Kurtosis	0.7	
Padé	0.9	3.4
Stretched exp.	1.2	4.5
Statistical	2.6	7.8
Fractional	4.6	7.8
Biexp.	3.6	4.7

The Kurtosis model is not included for b_{max} = 3000 s/mm^2.

appreciable signal do not strongly affect the overall median signal. To examine the validity of this approach, we extended our analysis to include differential levels of noise-floor correction applied during the fitting. The resulting changes in AICc are, in reality, only comparing the residuals, given that all other parameters of AICc are kept constant; we interpret the fact that the AICc values "converge" to similar values for all models at the nominal noise-floor correction value (scaling factor $k = 1$) as an indication that our method for noise-floor correction is sound. One can see, however, that accuracy of the noise-floor correction will affect conclusions regarding which representations have the best AICc, and this should be borne in mind when making such comparisons.

The second aim of this study was to assess the performance of diffusion signal representations in breast lesion classification, and noise-floor correction is important here also. In terms of the stability of fitted parameters without noise-floor correction, the results also vary: Among the 11 fitted parameters with AUC > 0.9 in Table 5, the ADC from the monoexponential representation and f from the biexponential representation stand out as the most stable, with only small and systematic changes for benign and malignant lesions (see Supporting Information Table S3). For the monoexponential

model, this is unsurprising given that at b = 600 s/mm^2 the signal-to-noise-floor ratio is still high (Figure 3). For the other representations, the fitted parameters are quite sensitive to lack of noise-floor correction, in particular the stretched exponential and fractional calculus representations, which both overestimate S_0 compared to all other representations, including compared to the measured S_0, and even more so when not using noise-floor correction. This suggests that fitting these two representations would benefit from including more data points at low b-values. The relatively high AUC from S_0 is, in itself, notable and suggests that the role of proton density and T_2 has perhaps been underexplored. S_0 values will also be influenced by exact details of scanner and protocol used and so are challenging in multiscanner studies unless normalized.

The results from the stretched exponential and fractional calculus representations are also interesting from another point of view. As illustrated in Figure 4, these representations had almost identical residuals, indicating very similar curve shapes. This comes as no surprise given that these mathematical representations are very similar. However, in terms of classification accuracy, they performed differently, with D from the fractional calculus giving higher AUC than α and DDC from the stretched exponential. This indicates that the information carried by the signal curves is transferred into the fitted parameters in different ways by these 2 representations, and that the fractional calculus representation more efficiently transfers information related to differentiating between benign from malignant lesions in a single parameter (D).

In general, the classification accuracy of DWI for this cohort of breast lesions was extremely high. Although this does not necessarily extrapolate to wider cohorts, it is nonetheless suggestive of the value of DWI in providing discriminatory power. The conventional ADC alone from the monoexponential fit for b = 200 and 600 s/mm^2 reached AUC = 0.96. The performance of ADC is improved to a very high AUC of 0.99, using some of the non-Gaussian representations. However, it is difficult to properly evaluate whether the increased scan time and analysis complexity for the high b-value protocol is warranted from a purely clinical point of view for differentiation of benign and malignant lesions compared to the simple ADC. A standard trace-weighted diffusion protocol with 2 b-values can be performed in 1 to 3 minutes, whereas the protocol used in this study had a substantially longer scan time of 7:46 (for all b-values up to 3000 s/mm^2). It would be interesting to evaluate the clinical value of high b-value protocol in the setting of, for example, prediction of treatment response in neoadjuvant therapy.

When looking at the two aims combined, one interesting feature is that there is no direct agreement between the quality of fit and classification accuracy. To illustrate, the PE and stretched exponential (SE) representations are ranked highly
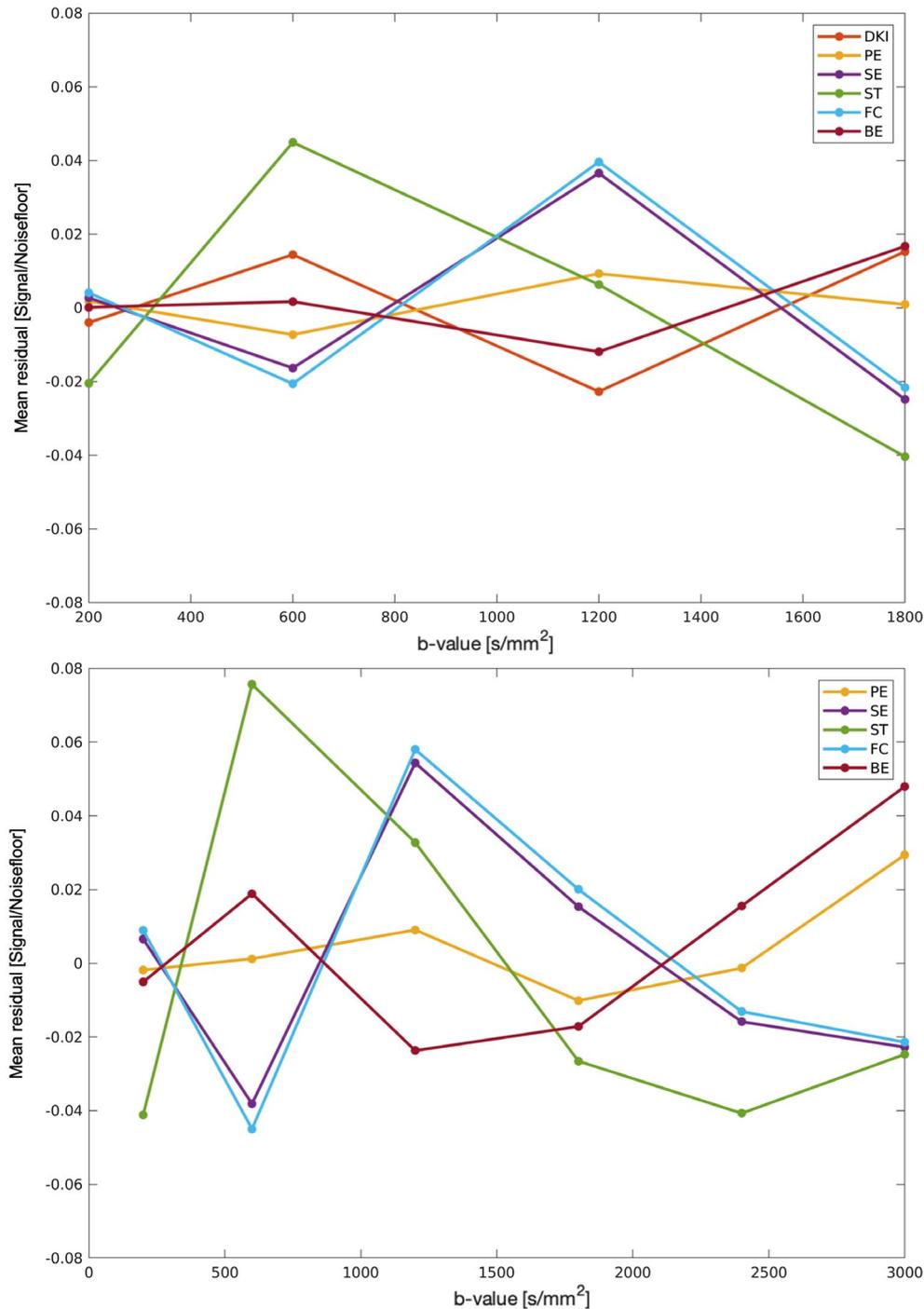
FIGURE 4 Mean residuals in units of signal-to-noise-floor at each b-value for each signal representation for the 2 different choices of b_{max}. BE, biexponential; DKI, diffusion kurtosis imaging; FC, fractional calculus; PE, Padé exponent; SE, stretched exponential; ST, statistical

in terms of overall quality of fit for both b_{max} conditions, but produce the lowest AUC (0.92 and 0.94 for PE and SE), and notably both perform worse than the ADC alone (AUC of 0.96).

We also report on the actual fitted values for the parameters of the different diffusion representations. To our knowledge, this represents the most comprehensive overview of observed

diffusion parameters in breast lesions across different models in the same data set. A study by Iima et al[30] reports mean K = 0.82 in malignant and K = 0.55 in benign lesions, whereas our study provides K = 1.01 and 0.39, respectively. For D_k, the values were 1.05 and 1.73 μm^2/ms in Iima et al[30] versus 1.14 and 1.92 μm^2/ms in our study. For both parameters, the mean values in Iima et al[30] are within the IQR of the values

TABLE 5 AUC values for classification of malignant/benign tumors for all of the extracted diffusion parameters, at different values of b_{max}

AUC of ROC		b_{max} [s/mm^2]		
		600	1800	3000
Monoexp.	S_0	0.858		
	ADC	**0.962**		
Kurtosis	S_0		0.851	
	K		**0.993**	
	D_k		0.942	
Padé	S_0		0.849	0.851
	δ_P		**0.920**	0.798
	D_p		0.918	**0.920**
Stretched exp.	S_0		0.776	0.776
	α		**0.938**	0.853
	DDC		0.838	0.76
Statistical	S_0		0.858	0.86
	σ		0.909	0.791
	ADC_{peak}		**0.973**	0.967
Fractional	S_0		0.776	0.78
	μ		0.687	0.733
	β		0.938	0.853
	D		0.984	**0.989**
Biexp.	S_0		0.847	0.851
	f		0.989	**0.998**
	D_1		0.511	0.511
	D_2		0.707	0.583

The highest parameter within all models are shown in bold.

in our study. It should be noted that the DWI protocols of the 2 studies are not identical.

To account for lesion heterogeneity, we have investigated possible differences in model parameters between the central and peripheral parts of the tumor (Supporting Information Tables S1 and S2). In particular, S_0 was systematically lower in the peripheral parts of the lesion, both for malignant and benign cases. This could originate from genuine differences in tissue property, like varying T_2, but also from the voxel composition (i.e., partial volume effects) and image acquisition process, like coil sensitivity profiles. In addition, for the malignant lesions, some diffusion coefficients were significantly higher in the periphery compared to the central parts of the ROI. This is also in agreement with a partial volume effect interpretation, where the central parts of the tumor have higher density of cancer cells with both lower diffusion coefficients and longer T_2. Further studies comparing DWI with histology are required to determine this.

The single voxel classification accuracy was evaluated (Supporting Information Table S4). The ranking of the various

non-Gaussian diffusion representations differed from the ROI mean analysis, indicating that sensitivity to measurement noise may vary among representations. One of the 4-parameter representations, the fractional calculus model, achieved the highest AUC. This is somewhat surprising, given that representations with higher degrees of freedom can be expected to have more local minima and thus be less robust in a low SNR situation. The relative increase of AUC from the monoexponential model to the best of the non-Gaussian representations are approximately the same as for the ROI mean analysis. Overall, the results indicate that the single voxel SNR in the reported acquisition protocol is sufficiently high to obtain robust fitting to non-Gaussian diffusion representations in the 200- to 3000-s/mm^2 range.

Similar analyses of different diffusion representations applied to brain tissue have concluded that the PE representation gave the best fit to brain data,[8] that the biexponential model was not appropriate for all parts of the brain, and that the signal can be described equally well with fewer parameters[7] using DKI. Our results for breast lesions are along the same lines. Regarding lesion differentiation using the biexponential representation, our study gives a different conclusion from Tamura et al, wherein no biexponential parameter was found to be a significant differentiator between malignant and benign lesions. On the contrary, we were able to extract a highly significant differentiator, the signal fraction f. Results in our study were comparable to DKI parameters and lesion differentiation accuracy obtained in Sun et al.[21] Furthermore, similar results to our study have been reported before by Iima et al,[30] where they observed perfect classification of breast lesions in a cohort of 26 patients using the kurtosis model in combination with IVIM. Note that in our study, we excluded $b = 0$ s/mm^2 from fitting; thus, these results might imply robustness of the DKI approach to the choice of the b-values.

A limiting factor is the available SNR in DWI. In the main part of this study, we averaged signal intensity across all voxels within the ROI before fitting to the different representations. This was done to improve the SNR of the data to enable differentiation between the quality of fit for the different representations. There is significant heterogeneity in tumors on the length scale relevant for MRI,[31] and this averaging procedure could potentially introduce additional partial volume effects not present on a single voxel level. This should be further investigated by comparing fitting results by averaging an increasing number of voxels. The comparison between central and peripheral ROI, as done in our study, is partially toward this end and shows some systematic differences, as already discussed. Another limitation of the study is that only single-parameter AUC was calculated, and not the combination (by logistic regression) of derived parameters within the model for evaluating the clinical potential of each model as a whole. Finally, a differently designed MR protocol could potentially result in different conclusions. For example,

acquiring 12 b-values in 3 directions instead of 6 b-values in 6 directions yields the same overall scan time, but if trace-weighted images are used for analysis, the degrees of freedom between these protocols are different and this would affect the AICc calculation. Choosing to use the raw images for each diffusion direction, as in our analysis, avoids this pitfall, and we thus believe our results to be less dependent on the specific combination of b-values versus directions, and should hold for other protocols with similar total scan time.

5 | CONCLUSION

In conclusion, the quality of fit for 6 diffusion signal representations were compared for the b-value range 200 to 3000 s/mm^2 for malignant and benign breast lesions. For noise-floor–corrected data, the PE representation gave the best fit overall (lowest residuals and AICc) with only 3 fitted parameters. All the models provided parameters with high AUC (>0.9) for the classification of benign and malignant lesions, with the highest AUC of 0.99 achieved by f (biexponential), K (kurtosis), and 0.989 for D (fractional calculus). The noise-floor correction level is affecting the fitting and the estimated parameters, and should thus be treated carefully. Finally, the supporting single voxel analysis showed that the SNR in this protocol, in the range of 10 to 20 at b = 0 s/mm^2, was sufficient to provide high classification accuracy for the statistical and fractional calculus diffusion representations, whereas the other non-Gaussian representations gave lower classification accuracy than the monoexponential model.

ORCID
Pål Erik Goa https://orcid.org/0000-0002-6866-6111

REFERENCES
1. Le Bihan D, Breton E. Imagerie de diffusion in-vivo par résonance magnétique nucléaire [Imagery of diffusion in vivo by nuclear magnetic resonance]. *Comptes-Rendus de l'Académie des Sciences*. 1985;93:27–34.
2. Cho GY, Moy L, Zhang JL, et al. Comparison of fitting methods and b-value sampling strategies for intravoxel incoherent motion in breast cancer. *Magn Reson Med*. 2015;74:1077–1085.
3. Jerome NP, d'Arcy JA, Feiweier T, et al. Extended T2-IVIM model for correction of TE dependence of pseudo-diffusion volume fraction in clinical diffusion-weighted magnetic resonance imaging. *Phys Med Biol*. 2016;61:N667–N680.
4. Le Bihan D, Breton E, Lallemand D, Aubin ML, Vignaud J, Laval-Jeantet M. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. *Radiology*. 1988;168:497–505.

5. Bihan DL. Apparent diffusion coefficient and beyond: what diffusion MR imaging can tell us about tissue structure. *Radiology*. 2013;268:318–322.
6. White NS, McDonald CR, Farid N, et al. Diffusion-weighted imaging in cancer: physical foundations and applications of restriction spectrum imaging. *Cancer Res*. 2014;74:4638–4652.
7. Kiselev VG, Il'yasov KA. Is the "biexponential diffusion" biexponential? *Magn Reson Med*. 2007;57:464–469.
8. Kristoffersen A. Optimized quantification of diffusional non-gaussianity in the human brain. *J Magn Reson Imaging*. 2013;38:1434–1444.
9. Jensen JH, Helpern JA, Ramani A, Lu H, Kaczynski K. Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging. *Magn Reson Med*. 2005;53:1432–1440.
10. Jensen JH, Helpern JA. MRI quantification of non-Gaussian water diffusion by kurtosis analysis. *NMR Biomed*. 2010;23:698–710.
11. Bennett KM, Schmainda KM, Bennett (Tong) R, Rowe DB, Lu H, Hyde JS. Characterization of continuously distributed cortical water diffusion rates with a stretched-exponential model. *Magn Reson Med*. 2003;50:727–734.
12. Niendorf T, Dijkhuizen RM, Norris DG, van Lookeren Campagne M, Nicolay K. Biexponential diffusion attenuation in various states of brain tissue: implications for diffusion-weighted imaging. *Magn Reson Med*. 1996;36:847–857.
13. Mulkern RV, Gudbjartsson H, Westin C-F, et al. Multi-component apparent diffusion coefficients in human brain. *NMR Biomed*. 1999;12:51–62.
14. Clark CA, Le Bihan D. Water diffusion compartmentation and anisotropy at high b values in the human brain. *Magn Reson Med*. 2000;44:852–859.
15. Schwarcz A, Bogner P, Meric P, et al. The existence of biexponential signal decay in magnetic resonance diffusion-weighted imaging appears to be independent of compartmentalization. *Magn Reson Med*. 2004;51:278–285.
16. Yablonskiy DA, Bretthorst GL, Ackerman JJ. Statistical model for diffusion attenuated MR signal. *Magn Reson Med*. 2003;50:664–669.
17. Zhou XJ, Gao Q, Abdullah O, Magin RL. Studies of anomalous diffusion in the human brain using fractional order calculus. *Magn Reson Med*. 2010;63:562–569.
18. Kärger J, Pfeifer H, Heink W. Principles and applications of self-diffusion measurements by nuclear magnetic resonance. *Adv Magn Reson*. 1988;12:1–89.
19. Novikov DS, Kiselev VG, Jespersen SN. On modeling. *Magn Reson Med*. 2018;79:3172–3193.
20. Tamura T, Usui S, Murakami S, et al. Biexponential signal attenuation analysis of diffusion-weighted imaging of breast. *Magn Reson Med Sci*. 2010;9:195–207.
21. Sun K, Chen X, Chai W, et al. Breast cancer: diffusion kurtosis MR imaging—diagnostic accuracy and correlation with clinical-pathologic factors. *Radiology*. 2015;277:46–55.
22. Ertas G, Onaygil C, Akin Y, Kaya H, Aribal E. Quantitative differentiation of breast lesions at 3T diffusion-weighted imaging (DWI) using the ratio of distributed diffusion coefficient (DDC). *J Magn Reson Imaging*. 2016;44:1633–1641.
23. Cox DR, Snell EJ. The choice of variables in observational studies. *Appl Statist*. 1974;23:51–59.

24. Vidić I, Egnell L, Jerome NP, et al. Support vector machine for breast cancer classification using diffusion-weighted MRI histogram features: preliminary study. *J Magn Reson Imaging*. 2018;47: 1205–1216.

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

26. Holland D, Kuperman JM, Dale AM. Efficient correction of inhomogeneous static magnetic field-induced distortion in echo planar imaging. *NeuroImage*. 2010;50:175–183.

27. Karunamuni RA, Kuperman J, Seibert TM, et al. Relationship between kurtosis and bi-exponential characterization of high b-value diffusion-weighted imaging: application to prostate cancer. *Acta Radiol*. 2018;59:1523–1529.

28. Köpf M, Corinth C, Haferkamp O, Nonnenmacher TF. Anomalous diffusion of water in biological tissues. *Biophys J*. 1996;70:2950–2958.

29. Hope TR, White NS, Kuperman J, et al. Demonstration of non-Gaussian restricted diffusion in tumor cells using diffusion time-dependent diffusion-weighted magnetic resonance imaging contrast. *Front Oncol*. 2016;6:179.

30. Iima M, Yano K, Kataoka M, et al. Quantitative non-Gaussian diffusion and intravoxel incoherent motion magnetic resonance imaging: differentiation of malignant and benign breast lesions. *Invest Radiol*. 2015;50:205–211.

31. Teruel JR, Heldahl MG, Goa PE, et al. Dynamic contrast-enhanced MRI texture analysis for pretreatment prediction of clinical and pathological response to neoadjuvant chemotherapy in patients with locally advanced breast cancer. *NMR Biomed*. 2014;27: 887–896.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

TABLE S1 Central ROI parameter values, median with inter-quartile region (IQR) for different values of b_{max}. S/NF = Signal-to-noise-floor

TABLE S2 Peripheral ROI parameter values, median with inter-quartile region (IQR) for different values of b_{max}. *Significant difference compared to Central ROI ($P < .05$). **Strongly significant difference compared to Central ROI ($P < .001$)

TABLE S3 Percentage difference in median values of fitted parameters between non-corrected and noise corrected data. Negative numbers mean lower parameter value in non-corrected data

TABLE S4 Single-voxel classification results. AUC of ROC values for classification of malignant and benign voxels pooled across all lesions. Values above 0.9 are shown in bold