Received: 11 April 2023 | Revised: 28 November 2023 | Accepted: 22 December 2023

DOI: 10.1002/mp.16935

RESEARCH ARTICLE

MEDICAL PHYSICS

Greybox: A hybrid algorithm for direct estimation of tracer kinetic parameters from undersampled DCE-MRI data

Aditya Rastogi[1,2] | **Phaneendra Kumar Yalavarthy**[1]

[1]Department of Computational and Data Sciences, Indian Institute of Science, Bangalore, India

[2]University Hospital Heidelberg, Heidelberg, Germany

Correspondence
Phaneendra Kumar Yalavarthy, Department of Computational and Data Sciences, Indian Institute of Science, Bangalore- 560012 India.
Email: phaneendra.k.yalavarthy@gmail.com

Funding information
SERB, Grant/Award Number: CRG/2018/000672; Prime Minister Research Fellowshop (PMRF)

Abstract

Background: A variety of deep learning-based and iterative approaches are available to predict Tracer Kinetic (TK) parameters from fully sampled or undersampled dynamic contrast-enhanced (DCE) MRI data. However, both the methods offer distinct benefits and drawbacks.

Purpose: To propose a hybrid algorithm (named as 'Greybox'), using both model- as well as DL-based, for solving a multi-parametric non-linear inverse problem of directly estimating TK parameters from undersampled DCE MRI data, which is invariant to undersampling rate.

Methods: The proposed algorithm was inspired by plug-and-play algorithms used for solving linear inverse imaging problems. This technique was tested for its effectiveness in solving the nonlinear ill-posed inverse problem of generating 3D TK parameter maps from four-dimensional (4D; Spatial + Temporal) retrospectively undersampled k-space data. The algorithm learns a deep learning-based prior using UNET to estimate the $\mathbf{K_{trans}}$ and $\mathbf{V_p}$ parameters based on the Patlak pharmacokinetic model, and this trained prior was utilized to estimate the TK parameter maps using an iterative gradient-based optimization scheme. Unlike the existing DL models, this network is invariant to the undersampling rate of the input data. The proposed method was compared with the total variation-based direct reconstruction technique on brain, breast, and prostate DCE-MRI datasets for various undersampling rates using the Radial Golden Angle (RGA) scheme. For the breast dataset, an indirect estimation using the Fast Composite Splitting algorithm was utilized for comparison. Undersampling rates of 8×, 12× and 20× were used for the experiments, and the results were compared using the PSNR and SSIM as metrics. For the breast dataset of 10 patients, data from four patients were utilized for training (1032 samples), two for validation (752 samples), and the entire volume of four patients for testing. Similarly, for the prostate dataset of 18 patients, 10 patients were utilized for training (720 samples), five for validation (216 samples), and the whole volume of three patients for testing. For the brain dataset of nineteen patients, ten patients were used for training (3152 samples), five for validation (1168 samples), and the whole volume of four patients for testing. Statistical tests were also conducted to assess the significance of the improvement in performance.

Results: The experiments showed that the proposed Greybox performs significantly better than other direct reconstruction methods. The proposed algorithm improved the estimated $\mathbf{K_{trans}}$ and $\mathbf{V_p}$ in terms of the peak signal-to-noise ratio by up to 3 dB compared to other standard reconstruction methods.

Conclusion: The proposed hybrid reconstruction algorithm, Greybox, can provide state-of-the-art performance in solving the nonlinear inverse problem of DCE-MRI. This is also the first of its kind to utilize convolutional neural network-

based encodings as part of the plug-and-play priors to improve the performance of the reconstruction algorithm.

KEYWORDS
AIF, compressive sensing, DCE-MRI, Fast-MRI, $\mathbf{K_{trans}}$, quantitative imaging, $\mathbf{V_p}$

1 | INTRODUCTION

Recent advances in medical imaging technology and tissue modeling have allowed clinicians to assess the tissue pathophysiological status for accurate diagnosis of diseases such as cancer, along with other clinically relevant action items such as prognosis, treatment planning, and assessment of treatment response. Techniques such as CT perfusion,[1] dynamic PET,[2] dynamic susceptibility contrast (DSC)[3] magnetic resonance imaging (MRI), arterial spin labelling (ASL),[4] diffusion-weighted imaging (DWI),[5] diffusion tensor imaging (DTI)[6] and dynamic contrast-enhanced imaging (DCE)[7] are popular quantitative imaging techniques. Apart from the first two, the rest are magnetic resonance (MR)-based techniques that do not subject the patient to ionizing radiation. In DCE-MRI, the bloodstream is injected with a T1 shortening contrast agent (CA),[8] and T1 weighted images of the region of interest (ROI) are periodically collected after a time interval. This leads to the use of four-dimensional (4D; 3D spatial + time) data acquisition to characterize healthy and unhealthy tissues. The principle behind DCE-MRI is that the vasculature in and around unhealthy tissue behaves differently from that of healthy tissue. This will cause different accumulation of CA in that region, and therefore different intensities in the dynamic scan.

Recent advances in medical imaging technology and tissue modeling have allowed clinicians to assess the tissue pathophysiological status for accurate diagnosis of diseases such as cancer, along with other clinically relevant action items such as prognosis, treatment planning, and assessment of treatment response. Techniques such as CT perfusion,[1] dynamic PET,[2] DSC[3] MRI, ASL,[4] DWI,[5] DTI[6] and DCE imaging [7] are popular quantitative imaging techniques. Apart from the first two, the rest are magnetic resonance (MR)-based techniques that do not subject patients to ionizing radiation. In DCE-MRI, the bloodstream is injected with a T1 shortening CA,[8] and T1 weighted images of the ROI are periodically collected after a time interval. This leads to the use of four-dimensional (4D; 3D spatial + time) data acquisition to characterize healthy and unhealthy tissues. The principle behind DCE-MRI is that the vasculature in and around unhealthy tissues behaves differently from that of healthy tissues. This will cause different accumulation of CA in that region, and therefore different intensities in the dynamic scan.

The Quantitative analysis of DCE-MRI can be performed using non-parametric (model-free) and parametric methods. Non-parametric methods analyze the tissue by characterizing the shape and size of the CA intensity versus time graph at each voxel in the ROI. Commonly used non-parametric indices include the peak enhancement (ΔS), wash-in,[9] wash-out[10] slope, time to peak enhancement (T_P), signal enhancement ratio (SER),[11] and area under the curve (AUC). However, although these indices are correlated with tissue activity, such as permeability and vascular density, they do not directly quantify these parameters. These parameters can be quantitatively assessed using pharmacokinetic (PK) models, the most popular being Patlak[12] and extended-Tofts or e-Tofts[13] models. These models are based on mathematical modeling of drug/agent exchange between the vasculature, cells, and extracellular extravascular space (EES). Patlak model characterizes tissues using vascular permeability ($\mathbf{K_{trans}}$) and plasma volume fraction ($\mathbf{V_p}$). The eTofts model uses an extra parameter to characterize the volume fraction of the EES. These parameters are also called tracer kinetic (TK) parameters because they characterize the dynamics of the tracer (CA) in the tissue. Clinically, $\mathbf{K_{trans}}$ parameter in DCE-MRI is important because it provides valuable information about tissue vascularity and perfusion, which has widespread applications in oncology,[13,14] treatment monitoring,[15] and drug development.[16] It helps clinicians and researchers better understand and manage various diseases and conditions by providing information on the vascular characteristics of the tissues. For example, one of the primary applications of DCE-MRI and $\mathbf{K_{trans}}$ is in oncology to assess tumour angiogenesis. tumour growth depends on angiogenesis, which involves the formation of new blood vessels. $\mathbf{K_{trans}}$ helps assess the rate at which CAs enter and exit the tumour tissue, providing valuable information on tumour vascularity. This information is crucial for the diagnosis, staging, and treatment planning of cancer. Although $\mathbf{K_{trans}}$ is a parameter of interest, $\mathbf{V_p}$ has clinical applications, such as quantification of blood volume and differentiation of tissue type.[17] It is the physiological equivalent of rCBV from DSC-MRI. $\mathbf{V_p}$ can aid in distinguishing different tissue types based on their vascular properties. For example, highly vascularized tumours may have a higher $\mathbf{V_p}$ than surrounding normal tissue, allowing for better tumour characterization. Moreover, similar to $\mathbf{K_{trans}}$,

$\mathbf{V_p}$ can be used to monitor the effectiveness of anti-angiogenic therapies and other treatments that affect tissue vascularity. Changes in $\mathbf{V_p}$ over time can provide information regarding the treatment response. Another important use for the accurate estimation of $\mathbf{V_p}$, and the most relevant in general, is to calculate $\mathbf{K_{trans}}$. Without prior knowledge of $\mathbf{V_p}$, the estimation of $\mathbf{K_{trans}}$ becomes a blind inverse problem.

To accurately estimate these TK parameters, high spatial and temporal resolution of the imaging data is important. However, spatial and temporal resolutions are inversely related, and in many applications, it is difficult to achieve Nyquist sampling simultaneously in the spatial and temporal dimensions. However, compressive sensing (CS) and parallel imaging-based techniques[18] have shown that sampling of k-t space below the Nyquist rate is possible, allowing simultaneous improvement of both spatial and temporal resolution, along with a reduction in scan time and patient motion artifacts. Algorithms for estimating TK parameters from undersampled k-t space can be categorized into indirect and direct estimation algorithms. Indirect estimation algorithms reconstruct anatomical images using the fundamentals of CS and use them to estimate the TK parameters. Examples include the works of Smith et al.,[19] Feng et al.[20] and Rosenkrantz et al.[21] Through tests on various animal and human scans, they demonstrated that undersampling rates (**R**) of 4-28.7× were possible without sacrificing diagnostic accuracy. They used $L_{1/2}$ or L_1 norms to induce sparsity with priors, such as wavelet transform or Total Variation (TV). Without recreating the anatomical pictures, direct estimation approaches enable TK parameters to be directly calculated from the undersampled k-t space. Guo et al.[22,23] and Dikaois et al.[24] have demonstrated that for higher undersampling rates, direct estimating strategies outperform indirect ones.

All methods mentioned above are conventional iterative techniques and require handcrafted prior to minimize the cost function. Researchers have recently used deep learning-based direct estimation techniques to estimate the TK parameters. Bliesener et al.[25] proposed a one-dimensional convolutional network to estimate the parameters at each pixel individually from fully sampled k-t data using both AIF and concentration maps as the network input. Cagdas et al.[26,27] used dilated convolution with a fully connected last layer to estimate TK parameters by treating the temporal dimension as channels without explicitly using AIF in the network. Kettelkamp et al.[28] improved this network by incorporating the AIF as an input. However, because the number of time samples was utilized as input channels, the network was not robust to data with different time samples. Moreover, as the PK model is learned implicitly, the network is also inflexible to the choice of PK model. This shortcoming can be addressed using indirect deep learning methods for TK parameter estimation using MR image reconstruction networks such as SpiNet,[29] MoDL[30] and ISTA-Net.[31] The advantage of this is that the hand crafted prior is learnt using a data driven approach. However, it was shown in [32] that conventional direct estimation techniques perform better than indirect deep-learning-based techniques for high undersampling rates.

This study presents the Greybox algorithm, a hybrid technique that combines the adaptability of iterative reconstruction models with respect to the number of time samples and the advantages of data-driven priors from PK models. The Greybox algorithm is an amalgamation of the conventional iterative estimation technique (Whitebox) and deep learning-based techniques (Blackbox). This algorithm was inspired by plug-and-play algorithms[33] used for solving linear inverse problems in imaging. Recently, many algorithms have used deep learning-based priors[29–31,34] in the plug-and-play framework for solving linear inverse problems. While the deep learning-based prior is a widely explored method in the domain of inverse problems, these methods are primarily explored in cases where the measurement operator is known. In this study, $\mathbf{K_{trans}}$, which is the most important parameter, was estimated from an unknown $\mathbf{V_p}$, which is part of the measurement operator, and vice versa. This makes the current problem a **blind nonlinear inverse problem**, where the joint estimation of both the parameters of interest and the forward function is required. Recently, Gan et al.[35] also proposed a denoising prior for blind image deconvolution. However, unlike the proposed Greybox, they solved a linear inverse problem using a prior trained on a larger dataset. Furthermore, existing methods that use deep learning-based priors[29–31,34] require separate priors for each subsampling rate, and generative prior-based methods require very large training data to train, making the existing methods intractable as the subsampling rate changes. The prior in the proposed method employs a single trained network for all undersampling rates for a wide range of $2-20\times$ undersampling. Therefore, it can be used to reconstruct an undersampling rate without explicitly training a prior for that undersampling rate (can be seen as more of unsupervised). Finally, another novelty of this work is to demonstrate that even though a network is trained on very few samples for the difficult task of estimating TK maps using undersampled data and performs poorly in terms of TK parameter estimation when trained in an end-to-end fashion, it helps in better estimation of parameters when used as a prior. This is unlike other plug-n-play-based algorithms, which demonstrate results using fine-tuned denoisers. Earlier work[36,37] showed that during training, the neural network (UNET in this study) prioritized learning low-frequency details first. In this study, the objective was to learn a prior for low-frequency details in the TK maps. This learned UNET prior is then utilized to regularize the error in the reconstruction of low-frequency components of the TK

parameter maps. In summary, the novelty of this work includes:

1. First algorithm to solve a non–linear blind inverse problem using a deep prior trained on very small datasets.
2. Trained prior can be utilized for multiple undersampling rates even those not seen during training.
3. Demonstration of reconstruction of TK parameter maps using an imperfect deep prior to regularize error in estimating low frequency components.

The results of the proposed approach were compared to those of a direct estimation method employing Total Variation (*TV*)[38] as carried out in Ref. [22]. A two-tailed t-test for unequal variance was utilized to assess if there was any significant difference in the performance of the two methods, wherein peak signal-to-noise ratio (PSNR) and Structural Similarity Index Measure (SSIM)[39] were utilized for comparison.

2 | BACKGROUND

2.1 | Notations

In the presented work, capital boldface letters such as **M** to represent matrices, and tiny boldface letters such as **v** to represent vectors. The vectors θ and $\hat{\theta}$ are used to define the 3D location in the image and frequency domains, respectively. The $p^{th} - norm$ of a vector **v** is represented by $||\mathbf{v}||_p$, which is defined as $||\mathbf{v}||_p = (\Sigma_n |v_n|^p)^{1/p}$, where v_n is the n^{th} element of the vector **v**. To distinguish between fully sampled and undersampled k-t space data, the tiny boldface symbols $k(\hat{\theta}, t)$ and $\hat{k}(\hat{\theta}, t)$ are used respectively.

2.2 | Image reconstruction

The Patlak model[12] was employed in this study as the PK model to quantify the exchange of CA between the vasculature and the EES. The rate at which CA permeates from the vasculature to the EES is represented by $\mathbf{K_{trans}}$ (min^{-1}), while $\mathbf{V_p}$ ($0 \leq \mathbf{V_p} \leq 1$) denotes the volume fraction of blood vessels in the tissue. The forward and inverse problems are delineated in detail in Ref [32].

As outlined in Section 1, algorithms for estimating TK parameters from undersampled k-t space can be categorized into indirect and direct estimation algorithms. Direct method of reconstruction uses compressive sensing (CS) schemes to directly estimate TK maps from undersampled k-t space data.[22] It was shown in Ref. [22] and [32] that direct reconstruction outperforms indirect reconstruction for TK parameter estimation specifically at higher undersampling rates. The optimization function for direct estimation of $\mathbf{K_{trans}}$ and $\mathbf{V_p}$ can be written as

$$\mathbf{K}^*_{\mathbf{trans}}(\theta), \mathbf{V}^*_{\mathbf{p}}(\theta) = \underset{\mathbf{K_{trans}}, \mathbf{V_p}}{\operatorname{argmin}} ||f(\mathbf{K_{trans}}(\theta), \mathbf{V_p}(\theta)) - \hat{\mathbf{k}}(\hat{\theta}, t)||_2^2 \tag{1}$$

$f(\cdot, \cdot)$ is a mapping from TK space to undersampled k-t space.

3 | METHODS

The direct iterative methods, while computationally slow, can be more robust than neural network-based methods. The performance of these methods is highly dependent on the proper prior selection. Based on the popularity of Plug and Play priors[33] for solving the linear inverse problems,[29,30,40] this paper suggests an approach called "Greybox" that uses deep learning to solve the non-linear inverse issue of TK parameter estimation. In this work, the discussion was limited to comparing direct estimation techniques using *TV* and deep learning-based regularization.

3.1 | Direct parametric estimation using *TV* regularization

The method for directly estimating parameters was thoroughly detailed in Ref. [22]. The anisotropic $||\cdot||_{TV}$ induces sparsity in $\mathbf{K_{trans}}$ and $\mathbf{V_p}$ maps due to the low permeability of CA in the majority of breast tissue. Thusly, the cost-function can be written as

$$\mathbf{K}^*_{\mathbf{trans}}(\theta), \mathbf{V}^*_{\mathbf{p}}(\theta) = \underset{\mathbf{K_{trans}}, \mathbf{V_p}}{\operatorname{argmin}} ||f(\mathbf{K_{trans}}(\theta), \mathbf{V_p}(\theta)) - \hat{\mathbf{k}}(\hat{\theta}, t)||_2^2 + \lambda_1 ||\mathbf{K_{trans}}(\theta)||_{TV} + \lambda_2 ||\mathbf{V_p}(\theta)||_{TV} \tag{2}$$

Similar to Ref. [22], this equation is solved alternatively by minimizing $\mathbf{K_{trans}}$ and $\mathbf{V_p}$ as shown in Equation (3) and Equation (4)

$$\mathbf{K}^*_{\mathbf{trans}}(\theta) = \underset{\mathbf{K_{trans}}}{\operatorname{argmin}} ||f(\mathbf{K_{trans}}(\theta), \mathbf{V_p}(\theta)) - \hat{\mathbf{k}}(\hat{\theta}, t)||_2^2 + \lambda_1 ||\mathbf{K_{trans}}(\theta)||_{TV} \tag{3}$$

$$\mathbf{V}^*_{\mathbf{p}}(\theta) = \underset{\mathbf{V_p}}{\operatorname{argmin}} ||f(\mathbf{K_{trans}}(\theta), \mathbf{V_p}(\theta)) - \hat{\mathbf{k}}(\hat{\theta}, t)||_2^2 + \lambda_2 ||\mathbf{V_p}(\theta)||_{TV} \tag{4}$$

The solutions to Equation (3) and Equation (4) are obtained using the limited memory Broyden–Fletcher–Goldfarb–Shanno (l-BFGS) algorithm.[41] The gradients were computed in accordance with the method outlined in Ref. [22] by relaxing the norm as described

in Ref. [42]. The TK parameters were also estimated without regularization (i.e., setting λ_1 and λ_2 to zero), resulting in *L*2 estimates. This procedure is also equivalent to solving Equation (1) in a cyclic manner. The algorithms described in this subsection are considered "Whitebox" methods, as each step of these algorithm is well understood.

3.2 | Proposed Greybox

In Greybox, a neural network is trained to estimate the prior knowledge necessary for promoting sparsity. The objective function for the Greybox algorithm can be expressed as follows (the dimension indications were removed for brevity):

$$J(\mathbf{K_{trans}}, \mathbf{V_p}) = ||f(\mathbf{K_{trans}}, \mathbf{V_p}) - \hat{\mathbf{k}}||_2^2 + \lambda_1||\Gamma_1(\mathbf{K_{trans}})||_2^2 + \lambda_2||\Gamma_2(\mathbf{V_p})||_2^2 \quad (5)$$

Equation (5) can be minimized by alternatively solving for $\mathbf{K_{trans}}$ and $\mathbf{V_p}$. The parameters have been estimated as shown below:

$$\mathbf{K}^{\mathbf{j}}_{\mathbf{trans}} = \underset{\mathbf{K_{trans}}}{\operatorname{argmin}} ||f(\mathbf{K_{trans}}, \mathbf{V}_\mathbf{p}^{\mathbf{j}-1}) - \hat{\mathbf{k}}||_2^2 + \lambda_1||\Gamma_1(\mathbf{K_{trans}})||_2^2 \quad (6)$$

$$\mathbf{V}^{\mathbf{j}}_{\mathbf{p}} = \underset{\mathbf{V_p}}{\operatorname{argmin}} ||f(\mathbf{K}^{\mathbf{j}}_{\mathbf{trans}}, \mathbf{V_p}) - \hat{\mathbf{k}}||_2^2 + \lambda_2||\Gamma_2(\mathbf{V_p})||_2^2 \quad (7)$$

Equation (6) was solved using Half Quadrature Splitting (HQS) by introducing an auxiliary variable *z*. The cost function corresponding to Equation (6) thus becomes

$$\hat{J}(\mathbf{K_{trans}}, z) = ||f(\mathbf{K_{trans}}, \mathbf{V}_\mathbf{p}^{\mathbf{j}-1}) - \hat{\mathbf{k}}||_2^2 + \lambda_1||\Gamma_1(z)||_2^2 + \mu||\mathbf{K_{trans}} - z||_2^2 \quad (8)$$

Equation (8) has been solved by keeping $\mathbf{K_{trans}}$ constant and solving for *z*, followed by solving for $\mathbf{K_{trans}}$ by keeping *z* constant as shown below.

$$z^j = \underset{z}{\operatorname{argmin}}\, \lambda_1||\Gamma_1(z)||_2^2 + \mu||\mathbf{K}^{\mathbf{j}-\mathbf{1}}_{\mathbf{trans}} - z||_2^2 \quad (9)$$

$$\mathbf{K}^{\mathbf{j}}_{\mathbf{trans}} = \underset{\mathbf{K_{trans}}}{\operatorname{argmin}} ||f(\mathbf{K_{trans}}, \mathbf{V}_\mathbf{p}^{\mathbf{j}-1}) - \hat{\mathbf{k}}||_2^2 + \mu||\mathbf{K_{trans}} - z^j||_2^2 \quad (10)$$

Similarly, Equation (7) was solved by introducing another auxiliary variable *y*. This simplifies Equation (7) to

$$y^j = \underset{z}{\operatorname{argmin}}\, \lambda_2||\Gamma_2(y)||_2^2 + \mu||\mathbf{V}_\mathbf{p}^{\mathbf{j}-\mathbf{1}} - y||_2^2 \quad (11)$$

$$\mathbf{V}^{\mathbf{j}}_{\mathbf{p}} = \underset{\mathbf{V_p}}{\operatorname{argmin}} ||f(\mathbf{K}^{\mathbf{j}}_{\mathbf{trans}}, \mathbf{V_p}) - \hat{\mathbf{k}}||_2^2 + \eta||\mathbf{V_p} - y^j||_2^2 \quad (12)$$

Equations (9) and (11) resembles denoising operation, and as such, can be replaced by any general-purpose denoising algorithm. In this study, two denoisers based on the UNET architecture[43] were employed as priors to address these equations. The training strategy for these UNETs is illustrated in Figure 1. The denoisers were trained using $\mathbf{K_{trans}}$ ($\mathbf{V_p}$) maps estimated from zero-filled undersampled k-t space, with undersampling rates (**R**) of 4×, 8×, 12×, 16× and 20×, as inputs and the map estimated from fully sampled data as ground truth. These trained networks were then used as denoising priors for $\mathbf{K_{trans}}$ and $\mathbf{V_p}$ parameters in $\Gamma_1(\mathbf{K_{trans}})$ and $\Gamma_2(\mathbf{V_p})$ respectively. The rationale behind employing maps estimated from multiple undersampling rates as input was that as the quality of the maps improves as one minimizes Equation (5), the maps will become increasingly similar to those estimated from lower undersampling rates. By utilizing inputs from multiple undersampling rates, a stable prior can be established at each iteration, and the algorithm can achieve better generalizability. The Greybox algorithm was visually depicted in the form of a flowchart in Figure 2.

4 | EXPERIMENTAL DETAILS

4.1 | Dataset

The DCE MRI datasets for the breast, prostate, and brain were used in the experiments conducted for this study. Below is a detailed description of these datasets.

4.2 | Breast dataset

For the majority of experiments performed in this study, the publicly available QIN Breast DCE-MRI dataset was utilized, which is sourced from The Cancer Imaging Archive (TCIA).[44] This dataset consists of breast DCE-MRI scans collected from a long-term clinical trial assessing the efficacy of neoadjuvant chemotherapy (NACT).[45] The scans were obtained at two time points: before the first treatment round (V1) and after the first round (V2). The scans were acquired using a Siemens 3T system with a four-channel bilateral phased array and a 3D gradient echo-based TWIST (Time-resolved angiography With Stochastic Trajectories) sequence for fat suppression. The flip angle, Time to echo (TE), and Time of repetition (TR) were set at 10°, 2.9 ms, and 6.2 ms, respectively, with a field of view (FOV) of 30–34 cm. The in-plane scan resolution was 320 × 320 with a slice thickness of 1.4 mm, and there were 28–32 anatomical volumes with a total scan time of 10 min, each containing 120–128 slices with a temporal resolution of 18–20 s. A gadolinium-based CA was administered at a dosage of 0.1 mmol/kg of body mass using a programmable injector, followed by a 20 mL saline flush at a rate of 2 mL/s. The breast data utilized in this study comprised V1 and V2 investigations from ten patients.

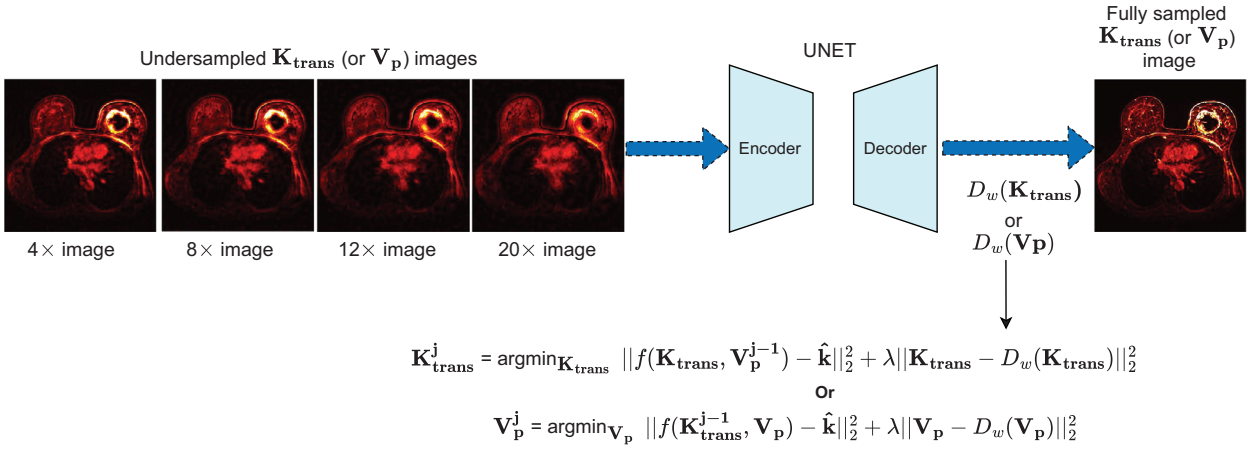

FIGURE 1 Network for learning prior for $\mathbf{K_{trans}}$ (and $\mathbf{V_p}$) reconstruction. Two separate UNETs (one for $\mathbf{K_{trans}}$ and another for $\mathbf{V_p}$) were trained to estimate the prior.

4.3 | Brain dataset

This manuscript also utilizes the DCE-MRI brain dataset, which is part of the RIDER (The Reference Image Database to Evaluate Therapy Response) NEURO MRI project published on TCIA.[46] The dataset comprises 19 individuals with recurrent glioblastoma who underwent multiple scans on a 1.5 T scanner. Each patient had two scanning sessions. To obtain dynamic images, 0.1 mmol/kg of Magnevist intravenous CA was injected into the patient's vein at a rate of 3 ccs/s, 24 s after the scan commenced. The 3D FLASH sequence was employed to capture the dynamic images, with flip angles of 25 degrees, TR of 3.8 ms, and TE of 1.8 ms. The voxel sizes and in-plane matrix resolution were 1 × 1 × 5mm and 256 × 256, respectively, and 16 axial slices were acquired. Out of 19 patients, 15 patients and their two sessions were used for training (12 for training and 3 for validating) the neural network. However, in the testing phase, only the first session scan of the four patients was used.

4.4 | Prostate dataset

The prostate dataset utilized in this study was obtained from the QIN Prostate database,[47] which is publicly available at TCIA.[44] The dataset comprised of DCE MRI of 22 patients, with each patient having only one scan session. Out of the 22 patients, 1 was excluded from the experiment as it consisted of only four time samples. Of the remaining 21, 15 patients were utilized for training the network, 3 for validation, and 3 for testing. Each scan consisted of 12–16 axial slices. The scans were conducted using a 3T scanner equipped with endorectal and phased array surface coils. The scanner employed a 3D SPGR sequence with flip angle set at 15 degrees, TE at 1.3 ms, and TR at 3.6 ms. The FOV of the scan was 26 square centimeters with a resolution of 1 mm^3 and an in-plane image size of 256 × 256 pixels. The volumetric scans, which were either 12 or 16 slices, were acquired at an interval of approximately 5 s. Prior to the administration of contrast, a protocol was followed which involved the acquisition of approximately five baseline scans to determine the baseline tissue parameters. Gadopentetate dimeglumine (0.15 mmol per kilogram) was then administered intravenously via a syringe pump at a rate of 3 mL/s, followed by a 20 mL saline flush at the same rate.

4.5 | Implementation

Using the Radial Golden Angle (RGA) undersampling pattern, Greybox training was conducted with undersampling rates (**R**) of 4×, 8×, 12×, 16× and 20×. To enhance robustness, random noise with a signal value of $\sigma = 0.01$ was added to the undersampled k-t space. A population-averaged Arterial Input Function (AIF) with a delay time specified by the patient's metadata was employed (delay time was the same for all patients in the prostate and brain datasets). The specifics of each algorithm's implementation are detailed below.

4.5.1 | Whitebox techniques

The Whitebox technique employs *L2* and *TV* based regularized reconstruction methods for estimating $\mathbf{K_{trans}}$ and $\mathbf{V_p}$. For the *L2* regularized reconstruction, the parameters λ_1 and λ_2 were kept as zero. In the case of the *TV* algorithm, the values of λ_1 and λ_2 were determined through adaptive grid search for each undersampling rate, and the optimal parameters were selected based on the highest average SSIM value, which was calculated using the formula $\text{SSIM}_{\text{avg}} = \frac{1}{1.5}\sum_{i=1}^{N}(\text{SSIM}_{\mathbf{K_{trans}}} + 0.5 * \text{SSIM}_{\mathbf{V_p}})_i$. These values were held constant for all patients. The initial

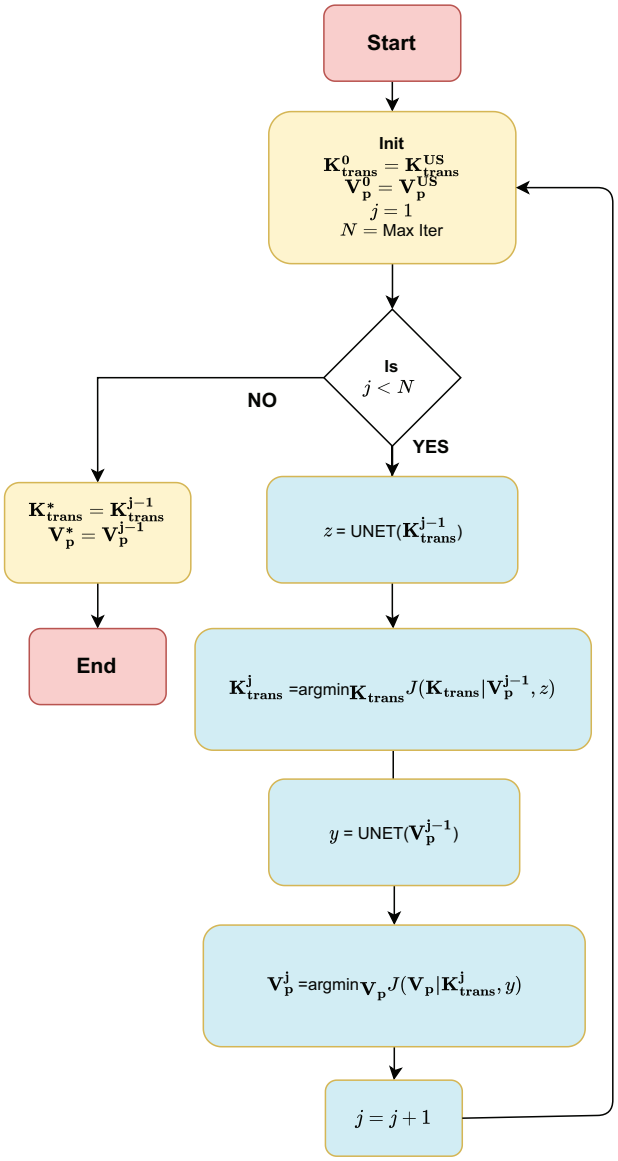

FIGURE 2 This flowchart illustrates the major steps of the proposed Greybox. The previous estimate of $\mathbf{K_{trans}}$ & $\mathbf{V_p}$ are used to calculate priors by passing them through their respective UNETs and are also used as initial estimates for next cycle of optimization of cost function. $\mathbf{K^0_{trans}}$ & $\mathbf{V^0_p}$ are the maps estimated from zero-filled undersampled $\mathbf{k-t}$ data.

guess for the *L2* and *TV* reconstruction was the tracker kinetic map estimated from the zero-filled undersampled data. The number of cyclic iterations was set to ten for the breast dataset and four for the brain and prostate datasets, as no significant improvement in results was observed beyond these values.

4.5.2 | Greybox techniques

The Greybox algorithm consists of two parts, (1) UNET based denoising prior and (2) steepest gradient based iterative minimizer. Two UNETs were used for denoising

TABLE 1 Hyperparameters of UNETs trained for prior estimate of $\mathbf{K_{trans}}$, $\mathbf{V_p}$ maps.

Hyperparameters	$\mathbf{K_{trans}}$	$\mathbf{V_p}$
Depth	4	4
Starting filters	32	16
Patch size	64	128
Downsampling rate	2.0	2.0
Filter increase rate	2.0	2.0
Dropout	0	0.1

Note: These hyperparameters were optimized empirically and were kept constant for breast, brain, and prostate datasets.

$\mathbf{K_{trans}}$ and $\mathbf{V_p}$ separately and as an estimator of prior. The network specifications for the $\mathbf{K_{trans}}$ and $\mathbf{V_p}$ UNETs are provided in Table 1. The mean squared error was employed as the loss function, and optimization was performed using the ADAM[48] optimizer with default settings and a learning rate of 5×10^{-4} for both networks. The networks were trained for 500 epochs with a batchsize of four. The input-label pair for training the $\mathbf{K_{trans}}$ network were $\mathcal{T} = \{\mathbf{K_{trans}}, \mathbf{K_{trans}}^{GT}\}_{i=1}^{M}$ where $\mathbf{K_{trans}}$ stands for $\mathbf{K_{trans}}$ map estimated from the zero-filled undersampled **k-t** data, $\mathbf{K_{trans}}^{GT}$ for the ground truth map estimated from fully sampled **k-t** data and M is the number of training samples. Similarly, for training the $\mathbf{V_p}$ UNET the input-label pairs were $\mathcal{T} = \{\mathbf{V_p}, \mathbf{V_p}^{GT}\}_{i=1}^{M}$. For the breast dataset comprising 10 patients, data from 4 patients were utilized for training, 2 for validation, and the entire volume of 4 patients for testing. A total of 1032 samples were extracted from the 4 training patients, and 752 samples were extracted from the 2 validation patients. For the prostate dataset consisting of 18 patients, 720 samples were extracted from 10 patients for training, 216 samples from 5 patients for validation, and the entire 4D volume of 3 patients for testing. For the brain dataset containing 19 patients, 3152 samples from 10 patients were utilized for training, 1168 samples from 5 patients for validation, and the complete volume of 4 patients for testing. In the implementation of the iterative minimizer (2), the regularization parameter selection was similar to that of the whitebox technique. Five iterations of the steepest descent method were performed, and the number of cyclic descent iterations was kept the same as those of the *L2* and *TV* regularized algorithms. Similar to the *TV* method, the parameters μ and η were estimated using adaptive grid search, and the Greybox algorithm was initialized with tracker kinetic maps estimated from zero-filled undersampled data.

4.5.3 | UNET based prior estimation

An examination of the performance of iterative reconstruction algorithms (including both Whitebox and

Greybox methods) in conjunction with two separate UNET-based denoising networks, which served as denoising priors, was conducted for the estimation of $\mathbf{K_{trans}}$ and $\mathbf{V_p}$. The training process for this analysis is detailed in Subsubsection 4.5.2. The results of this network are denoted as "UNET" in Section 5.

4.6 | Figures of merit

In this work, two figures of merit (metrics), namely PSNR and SSIM,[39] were used to compare the performance of the aforementioned approaches objectively. For both metrics, higher values indicate better reconstruction. The PSNR ($\in (-\infty, \infty)$) is measured in dB and is the log ratio of peak signal to reconstruction error for $\mathbf{K_{trans}}$ or $\mathbf{V_p}$. Whereas, the SSIM ($\in [-1, 1]$) is a unitless metric which represents the similarity of morphological aspects of the reconstructed image with the ground truth. A value closer to 1 denotes a higher performance of the reconstruction method

4.7 | Statistical test

The results from the Breast dataset of four patients were subjected to a two-tailed *t*-test for unequal variance, also known as the Welch test, for statistical analysis. The null hypothesis stated that there is no significant difference in the mean performance of the suggested Greybox and the *TV* reconstruction algorithm with respect to a particular metric. The Welch test was performed for the averaged PSNR and SSIM values for $\mathbf{R} = 8\times, 12\times$, and $20\times$ with a significance level of $\alpha = 0.05$. The results for $\mathbf{K_{trans}}$ and $\mathbf{V_p}$ are presented in Table 6 and Table 7, respectively. The probability that the performance gap between these two algorithms is due to chance is denoted by the symbol p, and a value of ≤ 0.05 indicates a statistically significant difference in their performance.

4.8 | Computational implementation

A Linux workstation with an Intel i9 CPU running at 2.10 GHz, 128 GB of RAM, and an NVIDIA Quadro RTX 8000 GPU with 48 GB of memory was used to do the calculations. While TensorFlow v1.13 was used to create the neural networks, MATLAB was used for the pre- and post-processing stages. With The *L2* and *TV* algorithm was implemented using MATLAB 2018b's parallel computing toolbox running on 12 threads. The code is made available as open-source for the suggested method's implementation at github.com/adityarastogi2k12/Greybox.

5 | RESULTS

5.1 | Breast data

This subsection evaluates the performance of the previously mentioned reconstruction algorithms on the breast data of four patients. To draw statistical inferences from the results, Welch's test was employed, as explained in Subsection 4.7. For visual comparison of the performance of all methods, the reconstructed parameter maps of a particular axial slice from one patient (patient 'A') were provided. Additionally, the Fast Composite Splitting Algorithm[49] (FCSA), a popular iterative algorithm for compressed MRI, was included in this experiment to estimate the $\mathbf{K_{trans}}$ and $\mathbf{V_p}$ for $8\times$, $12\times$, and $20\times$ undersampling rates. The FCSA algorithm reconstructed the MR image time series first and then estimated the TK parameters (indirect estimation). The iterative nature of the FCSA algorithm was chosen for comparison, as it shares one of the main advantages of the proposed Greybox method, namely, not requiring training for each undersampling rate unlike deep learning-based reconstruction methods. The regularizing parameters for the FCSA algorithm were also fine-tuned using adaptive grid search.

5.1.1 | Patient 'A'

Figure 3 and Figure 4 illustrate the maps of $\mathbf{K_{trans}}$ and $\mathbf{V_p}$ respectively, estimated using various methods and different undersampling rates. Specifically, the methods include I. Fully sampled data, II. zero-filled undersampled data, III. L^2 based direct reconstruction, IV. *TV* based direct reconstruction, V. FCSA based indirect estimation, VI. UNET trained as denoising prior, and VII. Proposed Greybox algorithm (**Grey**). The columns correspond to different undersampling rates, including $R = 8\times$, $R = 12\times$, and $R = 20\times$. Figure 3i and Figure 4i show the maps of $\mathbf{K_{trans}}$ and $\mathbf{V_p}$ in the whole breast region respectively, while Figure 3 ii and Figure 4 ii represent the corresponding maps in the ROI, marked by the green box. From Figure 3i, it can be observe that as $\mathbf{R}$ increases, the details in $\mathbf{K_{trans}}$ map estimated from undersampled data are lost. *L2*, *TV*, Indirect, UNET Prior and Greybox can better estimate the structure of the tumour; however, the high-frequency information in and around the tumour was lost in prior estimation. Figure 3 ii shows the ROI and it can be observed that *L2* and *TV* reconstructions show more hyperpermeability at inner and outer tumour boundaries as marked by the green arrow. This is biologically unlikely as large tumours have dead tissues at the centre, which have low permeability (as seen in the fully sampled image). The proposed Greybox algorithm has better visual fidelity to the ground truth and has preserved this detail better than

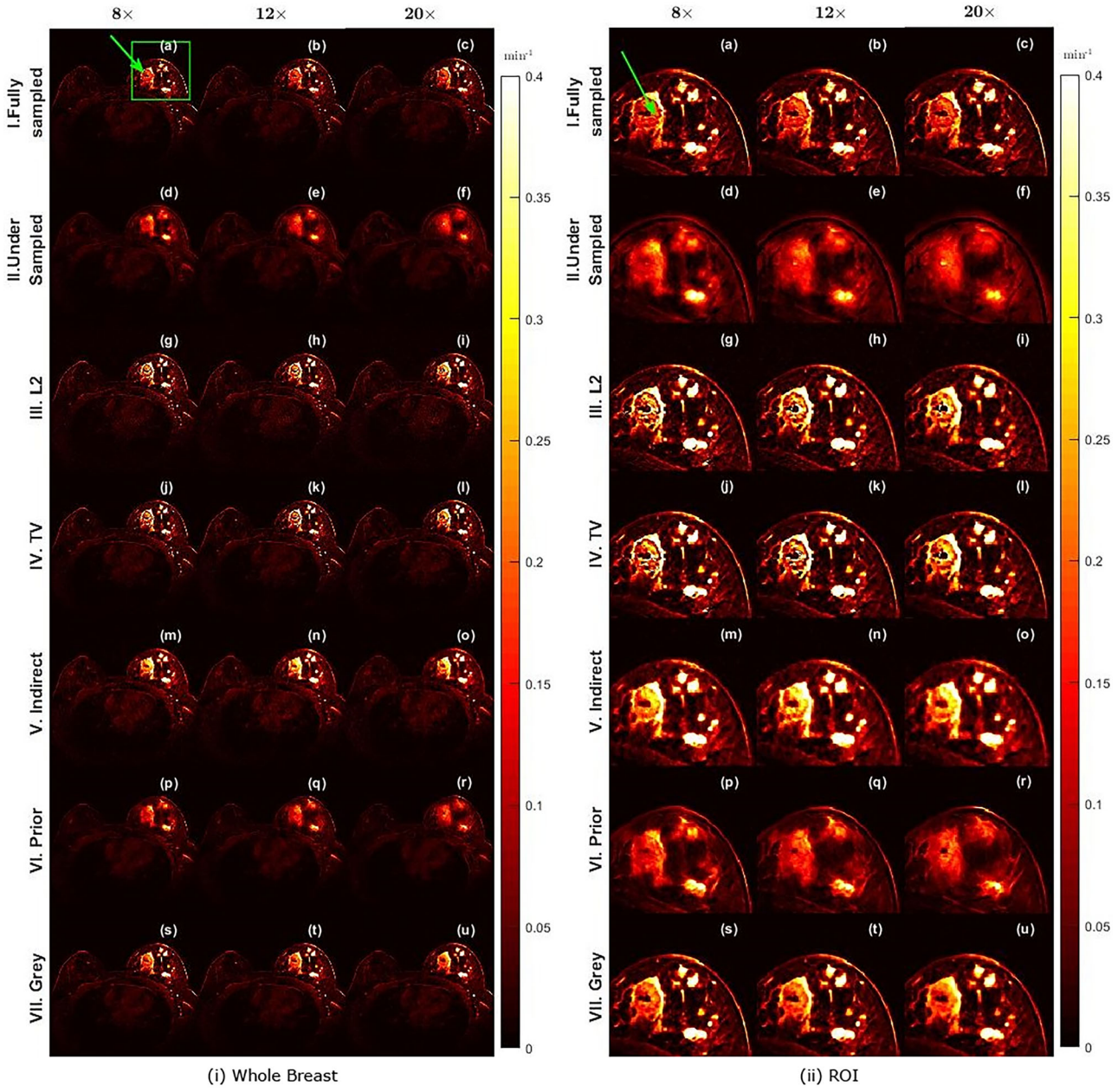

FIGURE 3 The figures display reconstructed $\mathbf{K_{trans}}$ breast data map of (i) representative axial slice and (ii) zoomed ROI (green-colored bounded box in (i)) of patient "A". The various under-sampling rates are mentioned at top of the columns and the reconstruction strategies are listed in front of the rows: I Ground truth, II. Undersampled data with zero paddings, III. *L*2 (no regularization) based direct reconstruction, IV. Direct *TV* regularized reconstruction, V. Indirect FCSA based reconstruction, VI. UNET based prior, VII. Proposed Greybox based direct reconstruction. From (i), it can be seen that as **R** increases, the details present in the map were lost, also for all **R**, *L*2, *TV* and Greybox perform better than denoising using neural network. However, from (ii), the *L*2 and *TV* reconstructions show hyperpermeability in the tumour region as marked by the green arrow. *L*2 and *TV* reconstructions also show hyperpermeability near the core of the tumour, which consists of dead tissue and generally has lower perfusion. The perfusion was better estimated by the Greybox algorithm [Digital copy in the full brightness of the screen is recommended for the viewing of these images]. FCSA, Fast Composite Splitting Algorithm; ROI, region of interest.

other methods. From Figure 4i, it can also be observed that as **R** increases, artifacts in $\mathbf{V_p}$ map estimated from undersampled data increases. *L*2, *TV*, Indirect, UNET prior and Greybox are able to reduce artifacts in the whole breast region. However, the UNET estimated map is not able to recover lost information like dead tissue at the core of the tumour. Figure 4 ii shows the ROI, and it can be observed that *L*2 and *TV* reconstructions better estimate the tumour map; however, the estimated map has specular noise in the whole image (as shown by the green arrow), with more noise in *L*2 reconstruction than *TV* reconstruction. As indicated by the cyan arrow, both algorithms also give a ring-like artifact around the tumour boundary. The proposed Greybox algorithm improved the reconstruction of the tumour structure without introducing the specular noise or ring artifact.

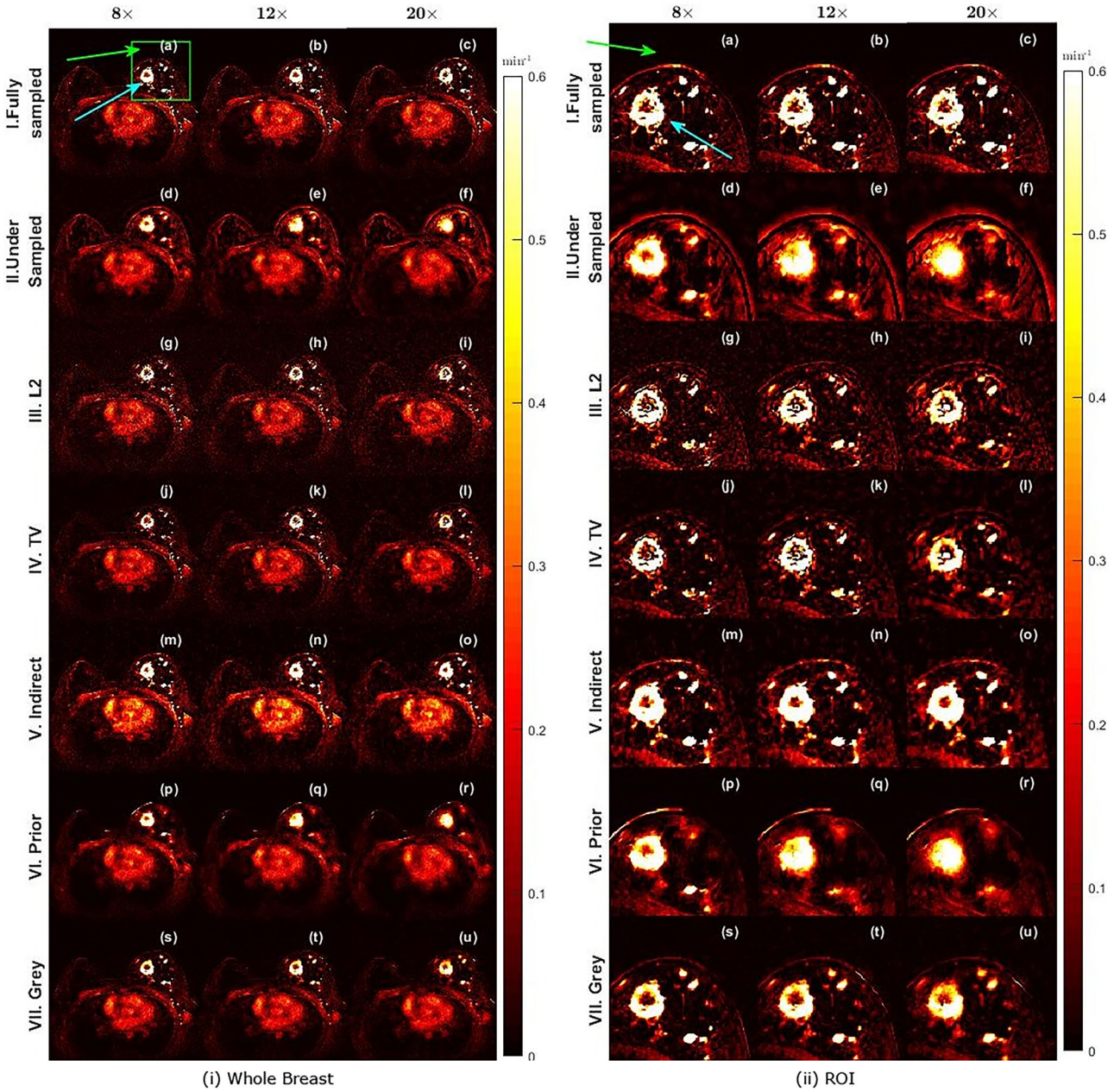

FIGURE 4 The figure displays the $\mathbf{V_p}$ reconstruction maps of Patient 'A' for the same representative slice as Figure 3. From (i), it can be inferred that as **R** increases, the details present in the map were lost, also for all **R**, *L2*, *TV*, Indirect and Greybox perform better than the prior. However, from (ii), the *L2* and *TV* reconstructions show halo-like artifacts on the tumour boundary as marked by the cyan arrow and specular noise in and outside the breast region (marked by green arrow). Similarly, $\mathbf{V_p}$ estimated from Indirect method also has specular noise outside the breast region and is also blurry. The $\mathbf{V_p}$ maps were better reconstructed by the Greybox algorithm without halo artifact at boundaries and less specular noise [Digital copy in the full brightness of the screen is recommended for the viewing of these images].

Figure 5 illustrates the absolute error maps computed by subtracting the $\mathbf{K_{trans}}$ parameter estimated by various reconstruction algorithms from the $\mathbf{K_{trans}}$ estimated from fully sampled data. It is evident from the figure that the error associated with the contrast-enhancing portion of the tumour is less in the prior estimate when compared to other methods (excluding Greybox). This is because the $\mathbf{K_{trans}}$ map denoised by the Prior network has effectively estimated low-frequency structures. However, the estimation of high-frequency structures is subpar. Conversely, the *L2* method demonstrates a superior ability to estimate high-frequency structures in comparison to the prior, and thus, the Greybox approach, which combines both these priors, exhibits superior estimation of $\mathbf{K_{trans}}$ with low error in both high-frequency and low-frequency regions. Figure 6 illustrates the absolute error maps for $\mathbf{V_p}$. Furthermore, it is apparent from the figure that the *TV*, *L2*, and FCSA-based Indirect

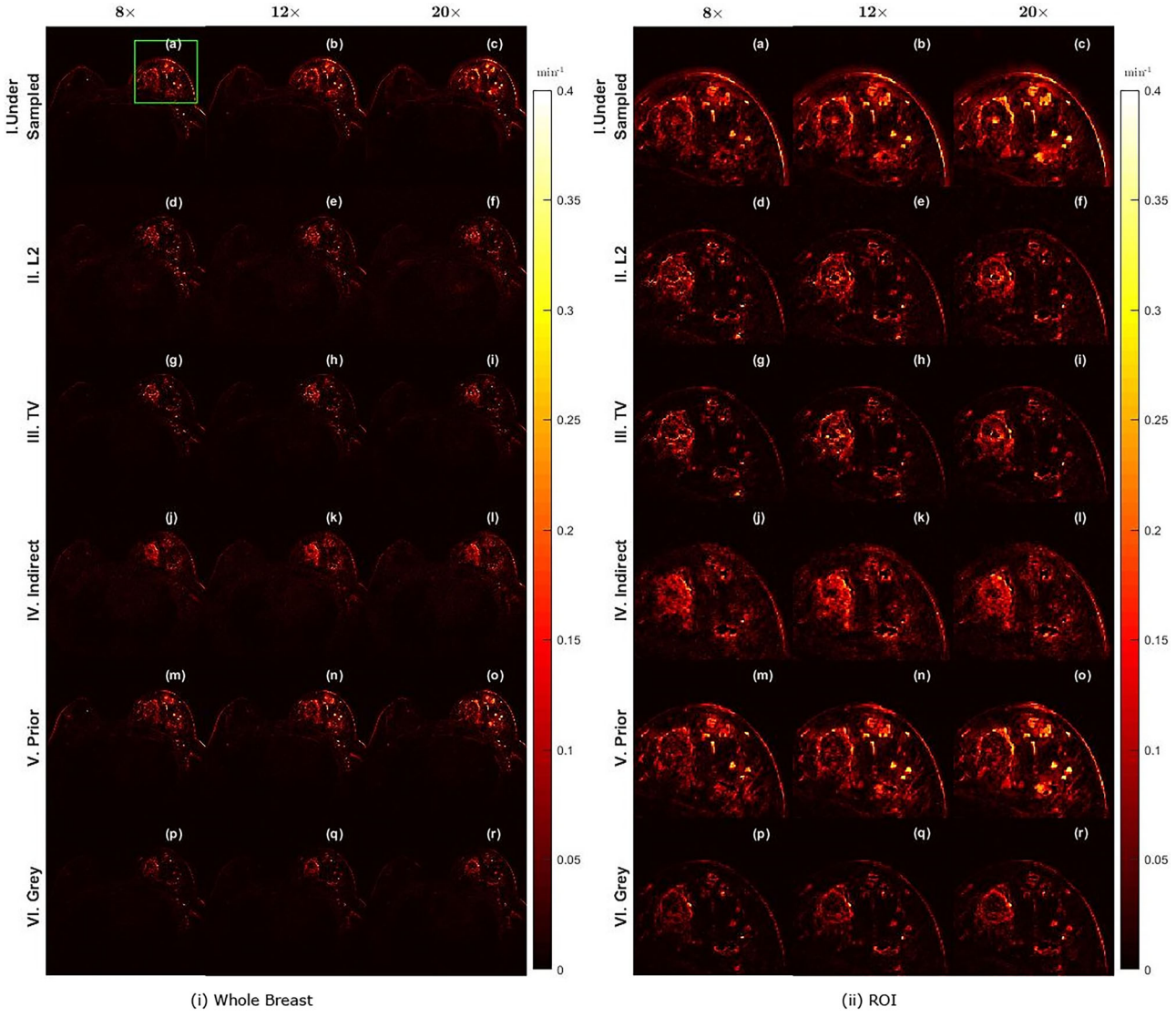

FIGURE 5 The figures display the error (with respect to fully sampled data) in reconstructed $\mathbf{K_{trans}}$ breast data map of (i) representative axial slice and (ii) zoomed ROI (green-colored bounded box in (i)) of patient "A". The various under-sampling rates are mentioned at the top of the columns and the reconstruction strategies are listed in front of the rows: I. Undersampled data with zero paddings, II. *L2* (no regularization) based direct reconstruction, III. Direct *TV* regularized reconstruction, IV. Indirect FCSA based reconstruction, V. UNET based prior, VI. Proposed Greybox based direct reconstruction. The images it can be seen that the Prior has less error in the Contrast-Enhancing tumour part i.e area with low spatial frequency. Therefore using this prior helps in better estimation of low frequency components in Greybox reconstruction. From the figure, it can be seen that Greybox estimation has the least error in estimating $\mathbf{K_{trans}}$ maps. [Digital copy in the full brightness of the screen is recommended for the viewing of these images].

techniques exhibit specular artifacts outside the breast region, which are absent in the Prior estimate while the prior displays a high error in the tumour region. The proposed Greybox displays comparatively lower error in both the tumour region and its surroundings.

5.1.2 | Average results

The performance of Greybox algorithm is compared with the performance of zero-filled reconstruction (**US**), *L2*, *TV*, FCSA and the prior estimated by UNET of $\mathbf{K_{trans}}$ and $\mathbf{V_p}$ on four patients for $\mathbf{R} = 8\times, 12\times$ and $20\times$. In Table 2 and Table 3, 3D PSNR and 3D SSIM are used as metrics for the comparison for $\mathbf{K_{trans}}$ and $\mathbf{V_p}$ respectively. Additionally, the mean and standard deviation of 2D PSNR and 2D SSIM, respectively, computed in the axial plane, were used as metrics for the comparison for $\mathbf{K_{trans}}$ and $\mathbf{V_p}$ in Table 4 and Table 5. A total of approximately 100 axial slices with visible tumours were selected from the testing phase of the four patients for the evaluation. From the tables, it can be seen that average 2D PSNR and SSIM are very close to 3D PSNR and SSIM in every case. The PSNR and SSIM values were also calculated in 2D because the standard deviation provides insights into the precision of the algorithm

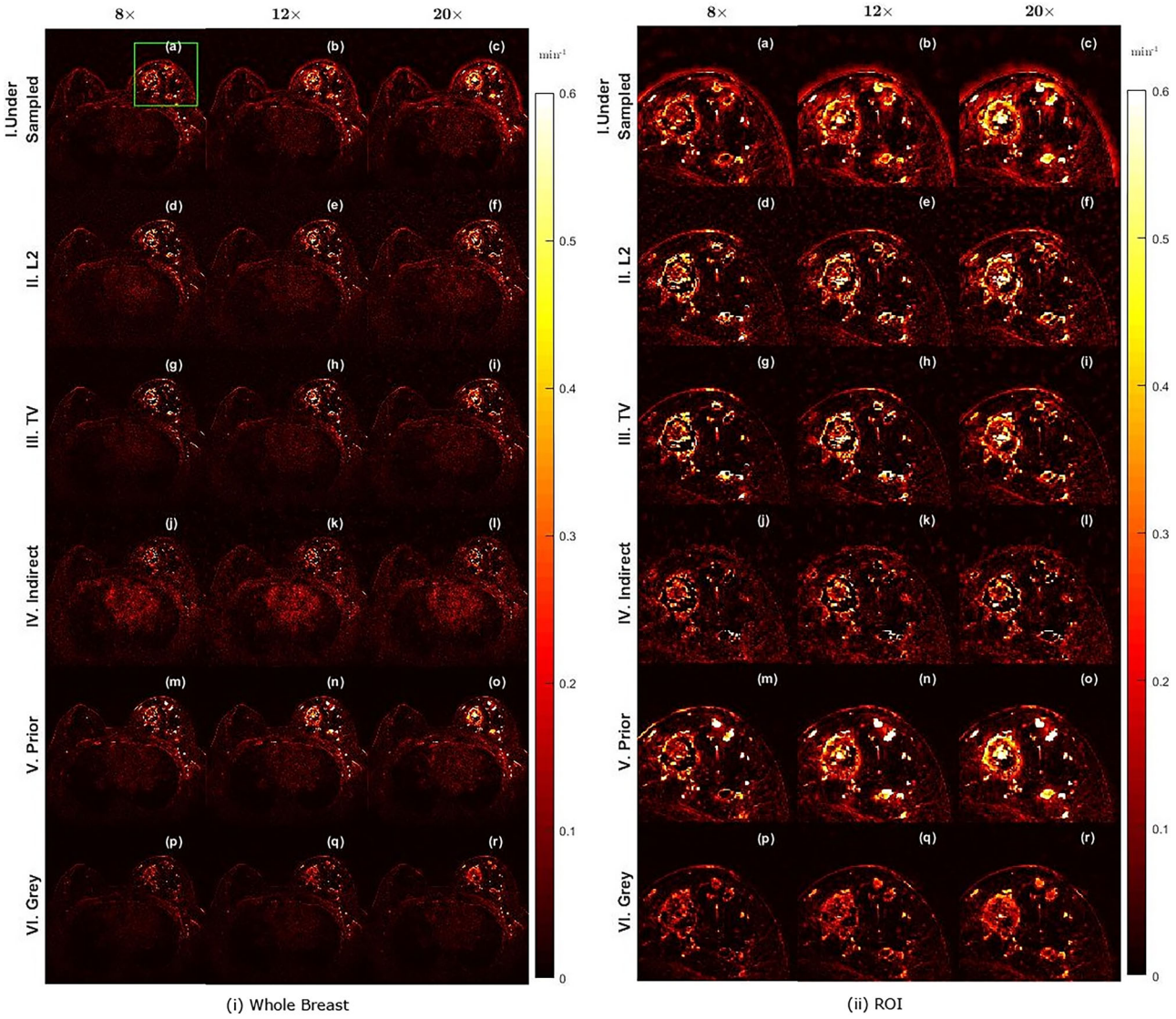

FIGURE 6 The figure displays the error (with respect to fully sampled data) in $\mathbf{V_p}$ reconstruction maps of Patient 'A' for the same representative slice as Figure 5. From the figure, it can be seen that Greybox reconstruction has less error outside the contrast-enhancing region. This can be attributed to the absence of high-frequency specular artifacts present in the Prior estimated by UNET. [Digital copy in the full brightness of the screen is recommended for the viewing of these images].

across different slices. Furthermore, the 2D estimation of PSNR and SSIM allowed for statistical analysis, which would not be possible otherwise. In both Table 4 and Table 5, it can be seen that the Greybox algorithm outperforms other algorithms in terms of both metrics. Moreover, based on Subsubsection 5.1.1, Table 4 and Table 5, it can be inferred that there is less improvement in the quality of $\mathbf{V_p}$ maps from algorithms like *L2, TV*, FCSA based indirect method and Greybox when compared to improvement in quality of $\mathbf{K_{trans}}$ map. However, the improvement by the Greybox algorithm is greater for $\mathbf{V_p}$ when compared to other algorithms. The experiment also demonstrates that the Greybox algorithm was capable of utilizing a network trained to denoise the TK parameter maps for estimating prior and improving the results, even though the network was trained on very small datasets and performed poorly.

5.1.3 | Statistical test

The Welch's test was conducted using PSNR and SSIM values of the $\mathbf{K_{trans}}$ and $\mathbf{V_p}$ maps estimated using the Greybox and *TV* algorithms for three different magnification levels (R = 8×, 12×, and 20×) on four patient breast data. The null hypothesis was that the average performance of the two algorithms in estimating the TK parameters was the same. The test is further explained in Subsection 4.7. A *p*-value less than 0.05 indicated a significant difference in the performance of the two

TABLE 2 This table shows the 3D (volume) computation of PSNR and SSIM for $\mathbf{K_{trans}}$ maps reconstructed using different reconstruction algorithms for different undersampling rates **R**.

Pat No.	Method	R = 8×		R = 12×		R = 20×	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
PAT A	US	34.44	0.9	33.49	0.88	32.54	0.86
	L2	36.08	0.89	35.69	0.88	35.66	0.87
	TV	38.3	0.97	37.79	0.96	37.02	0.94
	UNET Prior	35.59	0.93	34.52	0.92	33.47	0.9
	Indirect	37.81	0.93	37.02	0.92	36.36	0.92
	Grey	39.87	0.98	38.89	0.97	37.94	0.96
PAT B	US	36.99	0.91	36.04	0.89	34.99	0.86
	L2	37.84	0.92	38	0.91	37.79	0.9
	TV	39.01	0.96	39.01	0.95	38.84	0.94
	UNET Prior	37.74	0.93	37.01	0.92	36.3	0.91
	Indirect	38.88	0.94	37.93	0.93	37.39	0.93
	Grey	40.99	0.97	40.42	0.96	39.85	0.95
PAT C	US	35.29	0.91	34.37	0.89	33.32	0.86
	L2	36.67	0.92	36.37	0.9	35.93	0.89
	TV	37.04	0.94	36.71	0.93	36.24	0.91
	UNET Prior	35.57	0.93	34.65	0.91	33.65	0.89
	Indirect	36.12	0.92	35.36	0.91	34.43	0.89
	Grey	39.26	0.96	38.39	0.95	37.36	0.94
PAT D	US	35.29	0.91	34.37	0.89	33.32	0.86
	L2	36.67	0.92	36.37	0.9	35.93	0.89
	TV	37.04	0.94	36.71	0.93	36.24	0.91
	UNET Prior	35.57	0.93	34.65	0.91	33.65	0.89
	Indirect	36.12	0.92	35.36	0.91	34.43	0.89
	Grey	39.26	0.96	38.39	0.95	37.36	0.94

Note: The results are from the breast dataset of four patients used for testing. Abbreviations: PSNR, peak signal-to-noise ratio; SSIM, Structural Similarity Index.

TABLE 3 This table shows the 3D (volume) computation of PSNR and SSIM for $\mathbf{V_p}$ maps reconstructed using different reconstruction algorithms for different undersampling rates **R**.

Pat No.	Method	R = 8×		R = 12×		R = 20×	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
PAT A	US	25.78	0.65	24.7	0.61	23.63	0.57
	L2	26.35	0.59	25.81	0.55	24.95	0.51
	TV	27.65	0.76	27.22	0.71	26.25	0.63
	UNET Prior	26.3	0.75	25.4	0.73	24.68	0.7
	Indirect	27.83	0.72	27.71	0.7	27.2	0.68
	Grey	28.99	0.8	28.06	0.77	27.14	0.73
PAT B	US	24.64	0.63	23.67	0.6	22.59	0.55
	L2	25.64	0.61	25.12	0.55	24.72	0.52
	TV	26.45	0.73	26.02	0.68	25.48	0.63
	UNET Prior	25.51	0.76	24.75	0.74	24	0.71
	Indirect	24.94	0.69	24.27	0.67	23.88	0.66
	Grey	26.93	0.78	26.24	0.74	25.47	0.7
PAT C	US	24.07	0.65	23.06	0.6	21.82	0.55
	L2	24.57	0.59	24.04	0.55	23.3	0.51
	TV	24.95	0.7	24.55	0.65	23.81	0.59
	UNET Prior	24.41	0.73	23.52	0.7	22.49	0.67
	Indirect	24.09	0.66	23.5	0.64	22.83	0.61
	Grey	25.88	0.76	25.17	0.73	24.24	0.68
PAT D	US	25.69	0.6	25.02	0.58	24.22	0.54
	L2	26.05	0.56	25.86	0.54	25.03	0.49
	TV	28.37	0.73	27.9	0.69	27.16	0.63
	UNET Prior	27.32	0.7	26.78	0.68	26.13	0.66
	Indirect	23.53	0.64	22.21	0.59	21.56	0.58
	Grey	28.53	0.75	27.93	0.72	27.17	0.67

Note: The results are from the breast dataset of four patients used for testing. Abbreviations: PSNR, peak signal-to-noise ratio; SSIM, Structural Similarity Index.

methods. According to Table 6, Greybox outperformed the *TV* algorithm in terms of PSNR and SSIM for all four patients, except for patient C at R = 20× in the PSNR metric. Similarly, Table 7 shows that Greybox was superior to the *TV* algorithm for most cases.

5.2 | Prostate and brain data

To evaluate the effectiveness of Greybox across different datasets, the Greybox algorithm was trained and tested on the Prostate dataset for an undersampling rate of 8× and on the brain dataset for an undersampling rate of 20×. As an empirical rule in deep learning, more the training data, better is the performance of the network generally. As the data was limited, traditional 60-20-20 % split (approximately) was chosen for training, validation and testing data. Reducing the training or validation samples will lead to drop in performance due to two factors. Firstly, the diversity in training data will reduce and hence there are more chances that unseen structures can be present in test data. Secondly, having a smaller validation data might lead to bias in selecting the best network. This can happen if the validation dataset is not diverse enough. For example, if all the samples in validation dataset have large round tumours, then the selected network might not perform well on small, irregular tumours. From experiments performed on breast dataset and results tabulated in Table 4 and Table 5 it was inferred that the performance of *TV* was better than FCSA based indirect method in estimating both tracker kinetic parameters. Therefore, on Prostate and Brain dataest, FCSA was not used for comparison.

5.2.1 | Prostate data

For training the proposed Greybox algorithm, 10 patients were used as training data, 5 for validation and three for testing. The network was trained in the same manner as for breast data, and the algorithm was tested for **R** = 8×. Figures of merit, PSNR and SSIM, were utilized

TABLE 4 This table shows the mean and standard deviation of PSNR and SSIM for $\mathbf{K_{trans}}$ maps reconstructed using different reconstruction algorithms for different undersampling rates **R**.

		R = 8×		R = 12×		R = 20×	
Pat No.	Method	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
PAT A	US	34.54 ± 0.90	0.90 ± 0.01	33.58 ± 0.90	0.88 ± 0.01	32.63 ± 0.86	0.86 ± 0.01
	L2	36.12 ± 0.59	0.89 ± 0.02	35.74 ± 0.66	0.87 ± 0.02	35.73 ± 0.82	0.87 ± 0.02
	TV	38.37 ± 0.83	0.96 ± 0.01	37.85 ± 0.73	0.95 ± 0.01	37.09 ± 0.74	0.93 ± 0.01
	UNET Prior	35.70 ± 0.94	0.93 ± 0.01	34.61 ± 0.90	0.92 ± 0.01	33.54 ± 0.81	0.90 ± 0.01
	Indirect	37.88 ± 0.81	0.93 ± 0.01	37.08 ± 0.73	0.92 ± 0.01	36.44 ± 0.84	0.91 ± 0.01
	Grey	39.97 ± 0.92	0.97 ± 0.00	38.98 ± 0.92	0.97 ± 0.00	38.03 ± 0.91	0.95 ± 0.01
PAT B	US	37.02 ± 0.56	0.91 ± 0.01	36.08 ± 0.59	0.88 ± 0.01	35.03 ± 0.66	0.86 ± 0.01
	L2	37.93 ± 0.89	0.92 ± 0.02	38.06 ± 0.78	0.91 ± 0.02	37.85 ± 0.74	0.89 ± 0.01
	TV	39.07 ± 0.73	0.95 ± 0.01	39.08 ± 0.79	0.94 ± 0.01	38.92 ± 0.86	0.93 ± 0.01
	UNET Prior	37.78 ± 0.65	0.93 ± 0.01	37.06 ± 0.63	0.92 ± 0.01	36.35 ± 0.64	0.91 ± 0.01
	Indirect	39.22 ± 1.76	0.94 ± 0.02	38.37 ± 1.91	0.93 ± 0.02	37.85 ± 1.96	0.93 ± 0.02
	Grey	41.18 ± 1.32	0.96 ± 0.00	40.59 ± 1.22	0.96 ± 0.00	39.95 ± 0.98	0.95 ± 0.00
PAT C	US	35.33 ± 0.57	0.91 ± 0.01	34.40 ± 0.57	0.89 ± 0.01	33.36 ± 0.64	0.86 ± 0.01
	L2	36.42 ± 0.97	0.90 ± 0.02	36.03 ± 0.93	0.89 ± 0.02	35.99 ± 0.75	0.88 ± 0.01
	TV	37.12 ± 0.80	0.94 ± 0.01	36.79 ± 0.79	0.93 ± 0.01	36.32 ± 0.81	0.91 ± 0.01
	UNET Prior	35.62 ± 0.67	0.92 ± 0.00	34.70 ± 0.68	0.91 ± 0.00	33.71 ± 0.72	0.89 ± 0.01
	Indirect	36.23 ± 1.01	0.91 ± 0.01	35.43 ± 0.79	0.90 ± 0.01	34.53 ± 0.92	0.89 ± 0.01
	Grey	39.43 ± 0.64	0.96 ± 0.00	38.53 ± 0.65	0.94 ± 0.00	37.42 ± 0.75	0.94 ± 0.01
PAT D	US	35.19 ± 0.65	0.86 ± 0.01	34.56 ± 0.66	0.84 ± 0.01	33.96 ± 0.69	0.82 ± 0.01
	L2	39.24 ± 0.45	0.92 ± 0.00	38.72 ± 0.71	0.91 ± 0.01	38.21 ± 0.67	0.90 ± 0.01
	TV	41.98 ± 0.68	0.97 ± 0.00	41.42 ± 0.78	0.96 ± 0.00	40.75 ± 0.79	0.95 ± 0.01
	UNET Prior	36.20 ± 0.75	0.91 ± 0.01	35.65 ± 0.73	0.89 ± 0.01	35.17 ± 0.73	0.88 ± 0.01
	Indirect	36.00 ± 0.65	0.87 ± 0.01	34.78 ± 1.11	0.85 ± 0.02	33.90 ± 1.00	0.83 ± 0.02
	Grey	43.45 ± 0.66	0.97 ± 0.00	42.53 ± 0.69	0.97 ± 0.00	41.34 ± 0.75	0.96 ± 0.00

Note: The results are from the breast dataset of four patients used for testing.
Abbreviations: PSNR, peak signal-to-noise ratio; SSIM, Structural Similarity Index.

for objective comparison of performance and the results were tabulated in Table 8 for $\mathbf{K_{trans}}$ and Table 9 for $\mathbf{V_p}$. From the former, one can see that in terms of both PSNR and SSIM, the Greybox algorithm performs better than other methods and from *TV* by an average margin of ≈ 1dB (≈ 2dB for $\mathbf{V_p}$) in terms of PSNR and 0.03 for $\mathbf{V_p}$ points in terms of SSIM. In the case of estimation of $\mathbf{V_p}$, one can see from Table 9 that for patients 'B' and 'C', Greybox performed better than other methods in terms of both PSNR and SSIM. However, for patient 'A', UNET gave better results in terms of PSNR. Moreover, similar to the case of the breast dataset, the percentage recovery in $\mathbf{K_{trans}}$ estimate from any algorithm is more than $\mathbf{V_p}$.

5.2.2 | Brain data

For training Greybox algorithm on brain dataset, 10 patients data was used for training, 5 for validation and four for testing. The network was trained in the same manner as for breast data, and the algorithm was tested for **R** = 20×. Figures of merit, PSNR and SSIM, were utilized for objective comparison of performance and the results were tabulated in Table 10 for $\mathbf{K_{trans}}$ and Table 11 for $\mathbf{V_p}$. From the former, in terms of both PSNR and SSIM, the Greybox algorithm outperforms other methods, and from *TV* by an average margin of ≈ 2.5 dB in terms of PSNR and 0.03 points in terms of SSIM. Moreover, the standard deviation of Greybox based reconstruction is the least in terms of PSNR and SSIM among all the estimation algorithms (apart from undersampled reconstruction). In the case of estimation of $\mathbf{V_p}$, one can infer from Table 11 that the same trend as observed for $\mathbf{K_{trans}}$ holds good. The brain TK maps reconstructed using different reconstruction algorithms are visually illustrated in Figure 7 from three patients. Figure 7 i shows the reconstructed $\mathbf{K_{trans}}$ maps and Figure 7 ii shows the reconstructed $\mathbf{V_p}$ maps. By inspecting the images, it is inferred that *TV* and *L2* based direct reconstruction were corrupted by specular noise both in background and in the brain region. Moreover, these two techniques show hyperpermeability compared to ground truth. The UNET was not able to capture all the

TABLE 5 This table shows the mean and standard deviation of PSNR and SSIM for $\mathbf{V_p}$ maps reconstructed using different reconstruction algorithms for different undersampling rates **R**.

		R = 8×		R = 12×		R = 20×	
Pat No.	Method	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
PAT A	US	25.82 ± 0.56	0.64 ± 0.01	24.74 ± 0.54	0.60 ± 0.01	23.66 ± 0.54	0.57 ± 0.02
	L2	26.41 ± 0.72	0.59 ± 0.04	25.87 ± 0.79	0.55 ± 0.05	25.01 ± 0.72	0.51 ± 0.04
	TV	27.76 ± 1.00	0.75 ± 0.02	27.32 ± 0.96	0.70 ± 0.02	26.35 ± 0.99	0.62 ± 0.03
	UNET Prior	26.34 ± 0.65	0.74 ± 0.01	25.44 ± 0.60	0.72 ± 0.01	24.71 ± 0.57	0.70 ± 0.01
	Indirect	27.87 ± 0.61	0.70 ± 0.01	27.68 ± 0.63	0.68 ± 0.01	27.17 ± 0.70	0.66 ± 0.02
	Grey	29.03 ± 0.64	0.79 ± 0.01	28.11 ± 0.62	0.76 ± 0.01	27.18 ± 0.59	0.72 ± 0.01
PAT B	US	24.65 ± 0.31	0.63 ± 0.01	23.68 ± 0.33	0.59 ± 0.01	22.60 ± 0.34	0.55 ± 0.01
	L2	25.70 ± 0.74	0.60 ± 0.07	25.16 ± 0.62	0.55 ± 0.06	24.75 ± 0.53	0.52 ± 0.04
	TV	26.48 ± 0.49	0.72 ± 0.02	26.05 ± 0.45	0.67 ± 0.01	25.50 ± 0.46	0.62 ± 0.01
	UNET Prior	25.52 ± 0.31	0.75 ± 0.00	24.77 ± 0.41	0.73 ± 0.00	24.02 ± 0.43	0.70 ± 0.01
	Indirect	25.37 ± 1.98	0.68 ± 0.02	24.68 ± 1.93	0.66 ± 0.02	24.29 ± 1.93	0.65 ± 0.02
	Grey	26.95 ± 0.32	0.77 ± 0.01	26.25 ± 0.37	0.73 ± 0.01	25.48 ± 0.35	0.69 ± 0.01
PAT C	US	24.11 ± 0.59	0.64 ± 0.01	23.10 ± 0.56	0.59 ± 0.01	21.86 ± 0.58	0.54 ± 0.01
	L2	24.12 ± 0.61	0.55 ± 0.03	23.80 ± 0.66	0.52 ± 0.03	23.33 ± 0.54	0.51 ± 0.02
	TV	24.98 ± 0.56	0.68 ± 0.02	24.59 ± 0.54	0.64 ± 0.02	23.85 ± 0.61	0.58 ± 0.02
	UNET Prior	24.44 ± 0.54	0.72 ± 0.00	23.56 ± 0.59	0.69 ± 0.01	22.54 ± 0.68	0.66 ± 0.01
	Indirect	24.22 ± 1.05	0.65 ± 0.01	23.60 ± 0.91	0.62 ± 0.01	22.93 ± 0.94	0.60 ± 0.01
	Grey	25.90 ± 0.43	0.75 ± 0.01	25.23 ± 0.42	0.72 ± 0.01	24.26 ± 0.48	0.67 ± 0.01
PAT D	US	25.83 ± 1.22	0.60 ± 0.03	25.15 ± 1.20	0.57 ± 0.03	24.35 ± 1.20	0.54 ± 0.03
	L2	26.16 ± 1.14	0.55 ± 0.03	25.94 ± 0.88	0.53 ± 0.03	25.15 ± 1.13	0.48 ± 0.03
	TV	28.51 ± 1.27	0.72 ± 0.03	28.06 ± 1.31	0.68 ± 0.03	27.32 ± 1.36	0.62 ± 0.04
	UNET Prior	27.43 ± 1.05	0.70 ± 0.02	26.89 ± 1.10	0.68 ± 0.02	26.25 ± 1.14	0.65 ± 0.02
	Indirect	23.82 ± 1.84	0.62 ± 0.03	22.47 ± 1.66	0.58 ± 0.03	21.80 ± 1.56	0.56 ± 0.03
	Grey	28.64 ± 1.10	0.74 ± 0.02	28.04 ± 1.09	0.71 ± 0.02	27.29 ± 1.16	0.66 ± 0.03

Note: The results are from the breast dataset of four patients used for testing.
Abbreviations: PSNR, peak signal-to-noise ratio; SSIM, Structural Similarity Index.

TABLE 6 Results of the two-tailed Welch test on the PSNR and SSIM values of the *TV* and Greybox algorithm-reconstructed $\mathbf{K_{trans}}$ maps for various undersampling rates (**R**) for four patients' breast data.

		R = 8×				R = 12×				R = 20×			
Pat No.	Metric	TV	Grey	*p*-value	Sig?	TV	Grey	*p*-value	Sig?	TV	Grey	*p*-value	Sig?
PAT A	PSNR	38.37	39.97	$1.36e^{-09}$	Yes	37.85	38.98	$1.26e^{-06}$	Yes	37.09	38.03	$3.29e^{-05}$	Yes
	SSIM	0.96	0.97	$3.03e^{-10}$	Yes	0.95	0.97	$2.73e^{-11}$	Yes	0.93	0.95	$3.64e^{-11}$	Yes
PAT B	PSNR	39.07	41.18	$9.89e^{-08}$	Yes	39.08	40.59	$1.47e^{-05}$	Yes	38.92	39.95	0.000459	Yes
	SSIM	0.95	0.96	$3.37e^{-11}$	Yes	0.94	0.96	$1.52e^{-11}$	Yes	0.93	0.95	$1.99e^{-13}$	Yes
PAT C	PSNR	37.12	39.43	$3.64e^{-18}$	Yes	36.79	38.53	$2.32e^{-13}$	Yes	36.32	37.42	$6.02e^{-07}$	Yes
	SSIM	0.94	0.96	$1.63e^{-14}$	Yes	0.93	0.94	$1.69e^{-13}$	Yes	0.91	0.94	$1.01e^{-14}$	Yes
PAT D	PSNR	41.98	43.45	$1.11e^{-04}$	Yes	41.42	42.53	0.003	Yes	40.72	41.34	0.11	No
	SSIM	0.97	0.97	$2.24e^{-04}$	Yes	0.96	0.97	0.0006	Yes	0.95	0.96	0.01	Yes

Note: The probability that the performance gap between these two strategies was caused by chance is indicated here by the symbol p. If $p \leq 0.05$, the null hypothesis is rejected and there is a significant difference in the performance of both approaches. The "**Sig**?" column indicates whether or not the performance difference has been statistically significant.
Abbreviations: PSNR, peak signal-to-noise ratio; SSIM, Structural Similarity Index Measure.

TABLE 7 Results of the two-tailed Welch test on the PSNR and SSIM values of the *TV* and Greybox algorithm-reconstructed $\mathbf{V_p}$ maps for various undersampling rates (**R**) for four patients' breast data.

		R = 8×				R = 12×				R = 20×			
Pat No.	**Metric**	**TV**	**Grey**	***p*-value**	**Sig?**	**TV**	**Grey**	***p*-value**	**Sig?**	**TV**	**Grey**	***p*-value**	**Sig?**
PAT A	PSNR	27.76	29.03	$2.26e^{-07}$	Yes	27.32	28.11	$3.44e^{-04}$	Yes	26.35	27.18	$2.06e^{-04}$	Yes
	SSIM	0.75	0.79	$3.58e^{-14}$	Yes	0.70	0.76	$2.21e^{-18}$	Yes	0.62	0.72	$1.93e^{-21}$	Yes
PAT B	PSNR	26.48	26.95	$4.93e^{-04}$	Yes	26.05	26.25	0.10	No	25.50	25.48	0.892674	No
	SSIM	0.72	0.77	$1.65e^{-13}$	Yes	0.67	0.73	$4.60e^{-19}$	Yes	0.62	0.69	$2.78e^{-24}$	Yes
PAT C	PSNR	24.98	25.90	$1.46e^{-09}$	Yes	24.59	25.23	$2.69e^{-06}$	Yes	23.85	24.26	$4.56e^{-03}$	Yes
	SSIM	0.68	0.75	$1.84e^{-18}$	Yes	0.64	0.72	$4.88e^{-24}$	Yes	0.58	0.67	$1.35e^{-23}$	Yes
PAT D	PSNR	28.51	28.64	0.794	No	28.06	28.04	0.970	No	27.32	27.29	0.95	No
	SSIM	0.72	0.74	0.082	No	0.68	0.71	0.0181	Yes	0.62	0.66	$2.00e^{-03}$	Yes

Note: The probability that the performance gap between these two strategies was caused by chance is indicated here by the symbol p. If $p \leq 0.05$, the null hypothesis is rejected and there is a significant difference in the performance of both approaches. The "**Sig**?" column indicates whether or not the performance difference has been statistically significant.
Abbreviations: PSNR, peak signal-to-noise ratio; SSIM, Structural Similarity Index Measure.

TABLE 8 Mean and standard deviation of PSNR and SSIM for $\mathbf{K_{trans}}$ maps reconstructed using different reconstruction algorithms for an undersampling rate **R** = 20× on three patients prostate data.

		R = 8×	
Pat No.	**Method**	**PSNR**	**SSIM**
PAT A	**US**	37.55 ± 1.13	0.92 ± 0.02
	L2	39.99 ± 2.45	0.96 ± 0.02
	TV	40.76 ± 2.59	0.97 ± 0.01
	UNET	38.64 ± 1.95	0.94 ± 0.02
	Grey	41.85 ± 2.57	0.98 ± 0.01
PAT B	**US**	36.66 ± 0.71	0.90 ± 0.02
	L2	38.66 ± 2.37	0.95 ± 0.03
	TV	39.57 ± 2.46	0.96 ± 0.02
	UNET	37.43 ± 1.56	0.94 ± 0.02
	Grey	40.39 ± 1.96	0.97 ± 0.01
PAT C	**US**	37.34 ± 0.91	0.90 ± 0.02
	L2	40.84 ± 2.09	0.96 ± 0.01
	TV	41.99 ± 2.13	0.96 ± 0.01
	UNET	38.33 ± 1.22	0.93 ± 0.02
	Grey	42.42 ± 1.99	0.98 ± 0.01

Abbreviations: PSNR, peak signal-to-noise ratio; SSIM, Structural Similarity Index Measure.

TABLE 9 Mean and standard deviation of PSNR and SSIM for $\mathbf{V_p}$ maps reconstructed using different reconstruction algorithms for an undersampling rate **R** = 20× on three patients prostate data.

		R = 8×	
Pat No.	**Method**	**PSNR**	**SSIM**
PAT A	**US**	31.45 ± 1.78	0.80 ± 0.04
	L2	28.59 ± 2.31	0.82 ± 0.06
	TV	29.02 ± 2.44	0.86 ± 0.05
	UNET	32.41 ± 2.59	0.89 ± 0.03
	Grey	31.72 ± 2.40	0.88 ± 0.03
PAT B	**US**	29.32 ± 0.87	0.74 ± 0.02
	L2	29.26 ± 1.90	0.80 ± 0.04
	TV	29.98 ± 1.86	0.84 ± 0.04
	UNET	30.48 ± 0.94	0.88 ± 0.03
	Grey	31.97 ± 1.42	0.88 ± 0.04
PAT C	**US**	30.73 ± 1.55	0.73 ± 0.05
	L2	30.21 ± 2.49	0.82 ± 0.06
	TV	31.21 ± 2.59	0.85 ± 0.03
	UNET	32.02 ± 1.93	0.88 ± 0.03
	Grey	33.74 ± 2.15	0.89 ± 0.03

Abbreviations: PSNR, peak signal-to-noise ratio; SSIM, Structural Similarity Index Measure.

details during the reconstruction, however, using it as prior in the proposed Greybox algorithm improves the reconstruction performance.

Similar to Figure 5 and Figure 6, Figure 8 shows the error maps for $\mathbf{K_{trans}}$ parameter Figure 8i and $\mathbf{V_p}$ for Figure 8 ii estimate. From both subfigures, it can be seen that the proposed Greybox provides more accurate reconstruction.

5.3 | Runtime

For $\mathbf{K_{trans}}$ and $\mathbf{V_p}$ map estimate of a single axial slice, the runtime of several reconstruction strategies mentioned in this study were reported in Table 12 (rounded off to nearest second). The neural network reconstructs the TK maps the fastest, and the Greybox algorithm is the slowest, as can be seen from the table. This is

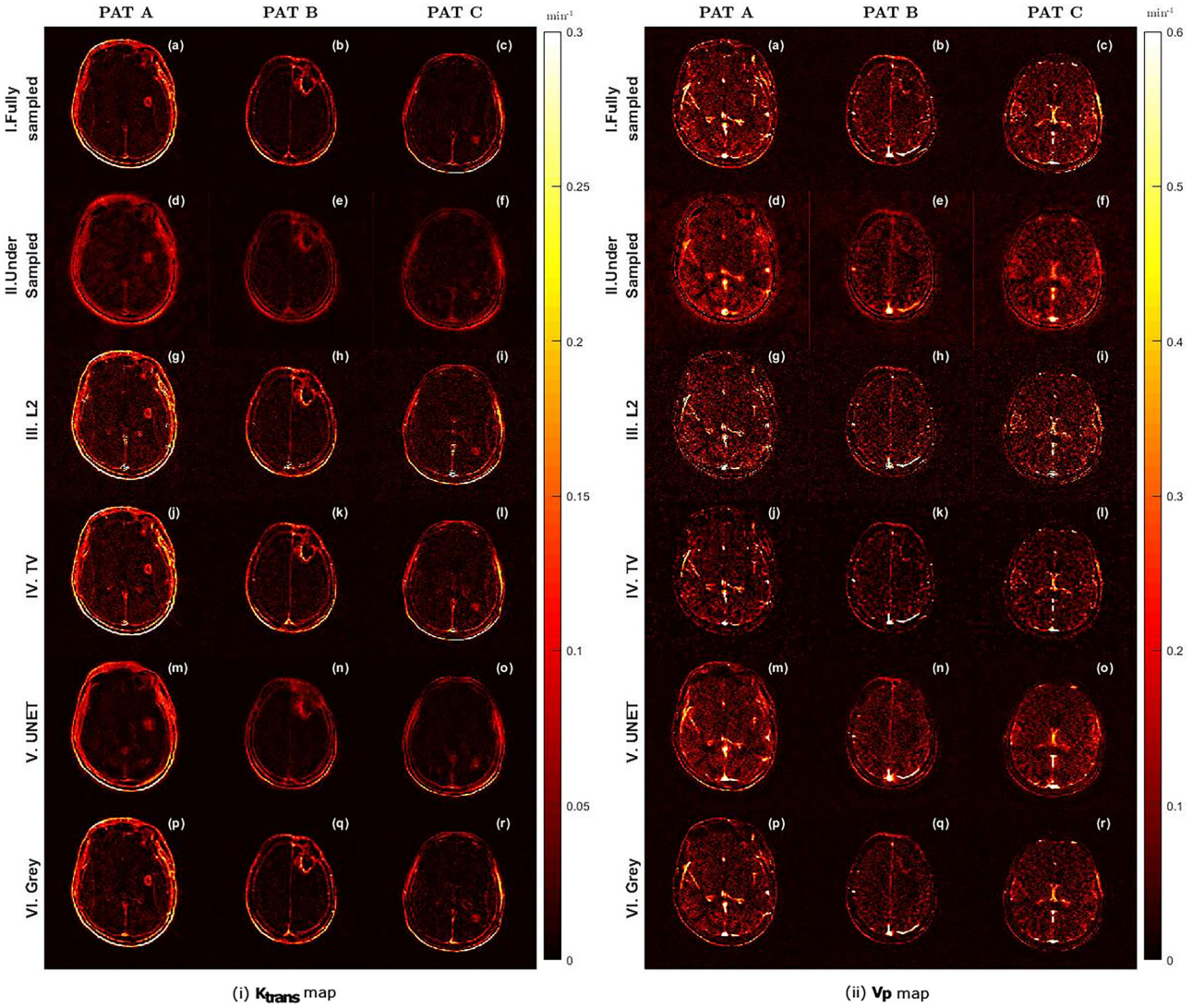

FIGURE 7 Reconstructed $\mathbf{K_{trans}}$ (i) & $\mathbf{V_p}$ (ii) maps for one representative slice each, of three **brain cancer patients** for $\mathbf{R} = 20\times$. The patient number was provided on top of each column. The reconstruction methods were indicated at the start of each row: I. Fully sampled data, II. Undersampled data with zero paddings, III. *L2* (no regularization) based direct reconstruction, IV. Direct *TV* regularized direct reconstruction, V. UNET based denoising prior, VI. Proposed Greybox based direct reconstruction. From (i), it is clear that both *TV* and *L2* reconstruction shows hyperpermeability in the $\mathbf{K_{trans}}$ maps and also suffer from specular noise. Moreover, from the $\mathbf{V_p}$ maps in (ii), it can be observed that the *TV* and *L2* reconstruction results have more background and in-brain specular noise than the proposed Greybox algorithm [Digital copy in the full brightness of the screen is recommended for the viewing of these images].

because each cycle of the Greybox algorithm requires loading the network twice (once for each parameter). More synchronized network integration with the iterative minimization approach can cut down on this time. Nevertheless, the gain in parameter estimate quality makes the runtime difference acceptable.

6 | DISCUSSION

The experimental results indicate that the performance of UNET when employed as a denoiser-based prior is not satisfactory in comparison to the *L2*, *TV*, or indirect FCSA-based parameter estimation. The reasons for this suboptimal performance are the complexity of the network's task and the limited size of the dataset. The task of the network is intricate, as it is required to estimate tracker kinetic parameters from input maps generated from undersampled data with a diverse range of degradation levels resulting from 2×–20× undersampling using only one set of trained parameters. The limited availability of samples for each organ presents a significant challenge in training the network. However, the results demonstrate that even with a poorly performing network, the parameter estimation can be significantly improved by utilizing the network as a prior. This method

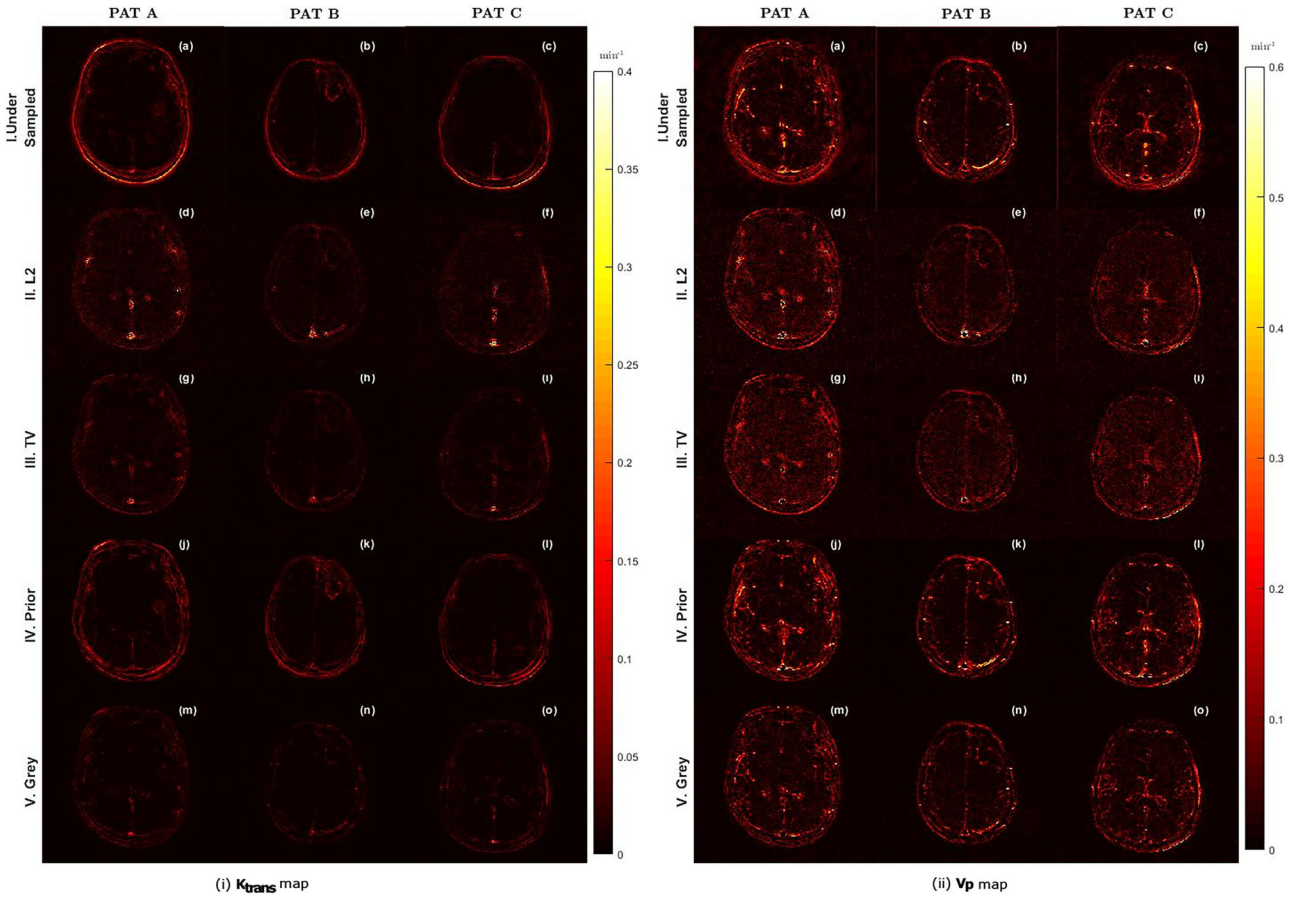

FIGURE 8 Figures shows the error maps for reconstructed $\mathbf{K_{trans}}$ (i) & $\mathbf{V_p}$ (ii) maps for one representative slice each, of three **brain cancer patients** for $\mathbf{R} = 20\times$. The patient number was provided on top of each column. The reconstruction methods were indicated at the start of each row: I. Undersampled data with zero paddings, II. *L2* (no regularization) based direct reconstruction, III. Direct *TV* regularized direct reconstruction, IV. UNET based denoising prior, V. Proposed Greybox based direct reconstruction. From both figures, it can be seen that Greybox reconstruction has the least error. [Digital copy in the full brightness of the screen is recommended for the viewing of these images].

not only surpassed the Indirect reconstruction based TK estimation but also direct *TV* based estimation. Figure 3 and Figure 5 show that while the UNET was unable to recover high-frequency components, it was able to better recover low-frequency smooth areas, which significantly improved performance when used as a prior. This is also demonstrated in experiments performed in Refs. [36, 37]. Therefore, using even an imperfectly trained UNET regularizer improves the reconstruction of low-frequency components. The reason direct estimation techniques perform better than indirect iterative techniques is that in indirect techniques, each image is reconstructed independently, and small inaccuracies in the reconstruction of individual images can cascade into larger cumulative errors at each step of further computations. In contrast, direct techniques directly optimize the objective function for TK parameters to minimize the cost. Another feature of the proposed algorithm is that unlike other networks for solving deep prior-based inverse problems, such as,[30,31,34] and,[29] which were also trained on small datasets, this algorithm is capable of reconstructing parameter maps for multiple undersampling rates in the range of 2–20× by solving a non-linear blind inverse problem.

The clinical utility of this algorithm is two-fold. First, it can be utilized in clinical scanners to directly estimate TK parameters from undersampled data. Through experimental results and existing literature,[22,23,32] it has been demonstrated that direct estimation of TK parameters is more accurate than indirect iterative estimation. The adoption of this technique in clinical practice has the potential to improve the accuracy of estimated maps. Secondly, the prior can be trained on a local in-house dataset at the clinical site, as the proposed Greybox algorithm has shown significant improvement in performance by training on a small dataset. Additionally, this algorithm can also be applied to other non-linear blind medical image reconstruction problems, such as Diffuse Optical Tomography[50] and Electrical Impedance Tomography.[51]

The present study also has inherent limitations. The most important limitation of this study is its retrospective

TABLE 10 Mean and standard deviation of PSNR and SSIM for $\mathbf{K_{trans}}$ maps reconstructed using different reconstruction algorithms for an undersampling rate **R** = 20× on four patients brain data.

		R = 20×	
Pat No.	Method	PSNR	SSIM
PAT A	US	31.33 ± 1.05	0.77 ± 0.07
	L2	31.11 ± 4.46	0.65 ± 0.25
	TV	36.62 ± 2.61	0.92 ± 0.04
	UNET	33.78 ± 1.66	0.89 ± 0.03
	Grey	39.01 ± 1.23	0.95 ± 0.01
PAT B	US	35.58 ± 2.00	0.87 ± 0.04
	L2	37.49 ± 1.40	0.88 ± 0.03
	TV	40.99 ± 0.89	0.94 ± 0.01
	UNET	37.19 ± 1.35	0.93 ± 0.02
	Grey	42.43 ± 0.92	0.96 ± 0.00
PAT C	US	34.07 ± 1.28	0.85 ± 0.04
	L2	32.75 ± 4.64	0.77 ± 0.13
	TV	37.04 ± 3.98	0.92 ± 0.03
	UNET	35.54 ± 2.07	0.91 ± 0.02
	Grey	39.87 ± 1.57	0.95 ± 0.00
PAT D	US	31.70 ± 1.01	0.80 ± 0.06
	L2	32.15 ± 3.55	0.68 ± 0.18
	TV	36.65 ± 2.21	0.91 ± 0.04
	UNET	34.61 ± 1.92	0.90 ± 0.03
	Grey	39.22 ± 0.86	0.95 ± 0.00

Abbreviations: PSNR, peak signal-to-noise ratio; SSIM, Structural Similarity Index Measure.

TABLE 11 Mean and standard deviation of PSNR and SSIM for $\mathbf{V_p}$ maps reconstructed using different reconstruction algorithms for an undersampling rate **R** = 20× on four patients brain data.

		R = 20×	
Pat No.	Method	PSNR	SSIM
PAT A	US	24.34 ± 1.41	0.64 ± 0.07
	L2	24.67 ± 2.03	0.63 ± 0.06
	TV	26.32 ± 1.60	0.63 ± 0.02
	UNET	25.11 ± 2.11	0.79 ± 0.05
	Grey	27.48 ± 1.49	0.78 ± 0.04
PAT B	US	27.67 ± 1.99	0.75 ± 0.07
	L2	28.73 ± 1.58	0.72 ± 0.04
	TV	29.95 ± 0.90	0.65 ± 0.01
	UNET	28.61 ± 1.05	0.84 ± 0.03
	Grey	30.83 ± 1.11	0.81 ± 0.03
PAT C	US	26.01 ± 1.64	0.73 ± 0.05
	L2	26.63 ± 2.59	0.68 ± 0.05
	TV	27.52 ± 1.88	0.65 ± 0.01
	UNET	26.18 ± 1.98	0.82 ± 0.02
	Grey	28.52 ± 1.38	0.78 ± 0.03
PAT D	US	24.17 ± 1.39	0.64 ± 0.08
	L2	25.08 ± 1.99	0.63 ± 0.06
	TV	26.46 ± 1.45	0.62 ± 0.02
	UNET	25.52 ± 1.60	0.80 ± 0.04
	Grey	27.85 ± 1.27	0.76 ± 0.04

Abbreviations: PSNR, peak signal-to-noise ratio; SSIM, Structural Similarity Index Measure.

TABLE 12 Reconstruction time for $\mathbf{K_{trans}}$ & $\mathbf{V_p}$ for a single slice data for the algorithms discussed in this work.

Method	Reconstruction time (s)
L2	62
TV	65
UNET	0.02
Grey	75

Note: The reported run times were rounded off to the nearest second.

nature. All experiments were performed on single coil simulation due to the unavailability of raw data. Whereas in clinics, multicoil acquisition is widely adapted. This limitation affects both the reconstruction performance and the highest undersampling rate tested in this study. The next limitation is that the prior is dataset dependent like most deep learning architectures. This will necessitate training a different model for different body organs. Another major limitation is that hyperparameters such as λ that were used to weight the regularizer must be tuned empirically. These regularization parameters can be trained using Deep Learning; however, it will restrict the number of iterations of the minimization algorithm, as is the case in many unrolled deep learning-based schemes.[29–31,34] Finally, the run-time of proposed Greybox is higher in its vanilla implementation, because the iterative reconstruction part of the algorithm runs on Matlab and the prior estimation part runs on Python. It can be significantly improved by utilizing C/C++ based libraries.

7 | CONCLUSION

This study presented a hybrid algorithm for the direct estimation of TK parameters $\mathbf{K_{trans}}$ and $\mathbf{V_p}$ from undersampled k-t data. The proposed algorithm, named "Greybox," solves a nonlinear inverse problem using a pre-trained neural network as a prior. The proposed Greybox algorithm was tested on a breast dataset with undersampling rates of 8×, 12×, and 20× on four patients, using PSNR and SSIM as metrics. The results showed that the proposed Greybox algorithm outperformed the existing direct reconstruction methods, as demonstrated by Welch's test on the reconstruction results from the data of four patients. The generalizability of the proposed Greybox algorithm was also tested on brain and prostate DCE-MRI data with undersampling rates of 20× and 8×, respectively, on four and three patients. The same trend of improved performance was observed in these cases. A detailed explanation of the

observations and limitations of the proposed approach was also presented.

ACKNOWLEDGMENTS

This work was supported by the Science & Engineering Research Board (SERB) core research grant (No. CRG/2018/000672). Aditya Rastogi acknowledges the Prime Minister Research Fellowshop (PMRF).

CONFLICT OF INTEREST STATEMENT

There are no conflicts of interest declared by Authors.

DATA AVAILABILITY STATEMENT

The breast, brain and prostate datasets were obtained from The Cancer Imaging Archieve (https://wiki.cancerimagingarchive.net/display/Public/Wiki). The code has been made available for enthusiastic users as open source at github.com/adityarastogi2k12/Greybox.

REFERENCES

1. Lui YW, Tang ER, Allmendinger AM, Spektor V. Evaluation of CT perfusion in the setting of cerebral ischemia: patterns and pitfalls. *AJNR Am J Neuroradioln*. 2010;31:1552-1563.
2. Rahmim A, Lodge MA, Karakatsanis NA, et al. Dynamic whole-body PET imaging: principles, potentials and applications. *Eur J Nucl Med Mol Imaging*. 2019;46:501-518.
3. Essig M, Shiroishi MS, Nguyen TB, et al. Perfusion MRI: the five most frequently asked technical questions. *AJR Am J Roentgenol*. 2013;200:24-34.
4. Alsop DC, Detre JA, Golay X, et al. Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: a consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. *Magn Reson Med*. 2015;73:102-116.
5. Baliyan V, Das CJ, Sharma R, Gupta AK. Diffusion weighted imaging: technique and applications. *World J Radiol*. 2016;8:785.
6. Lagopoulos J. Diffusion tensor imaging: an overview. *Acta Neuropsychiatr*. 2007;19:127-128.
7. Gribbestad IS, Gjesdal KI, Nilsen G, Lundgren S, Hjelstuen MHB, Jackson A. An introduction to dynamic contrast-enhanced MRI in oncology. In: *Dynamic contrast-enhanced magnetic resonance imaging in oncology*. Springer; 2005:1-22.
8. Yan Y, Sun X, Shen B. Contrast agents in dynamic contrast-enhanced magnetic resonance imaging. *Oncotarget*. 2017;8:43491.
9. Nagel E, Klein C, Paetsch I, et al. Magnetic resonance perfusion measurements for the noninvasive detection of coronary artery disease. *Circulation*. 2003;108:432-437.
10. Parker GJ, Suckling J, Tanner SF, Padhani AR, Husband JE, Leach MO. MRIW: parametric analysis software for contrast-enhanced dynamic MR imaging in cancer. *Radiographics*. 1998;18:497-506.
11. Hylton NM, Blume JD, Bernreuter WK, et al. Locally advanced breast cancer: MR imaging for prediction of response to neoadjuvant chemotherapy–results from ACRIN 6657/I-SPY TRIAL. *Radiology*. 2012;263:663-672.
12. Patlak CS, Blasberg RG, Fenstermacher JD. Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. *J Cereb Blood Flow Metab*. 1983;3:1-7. PMID: 6822610.
13. Tofts PS, Brix G, Buckley DL, et al. Estimating kinetic parameters from dynamic contrast-enhanced T1-weighted MRI of a diffusable tracer: standardized quantities and symbols. *J Magn Reson Imaging*. 1999;10:223-232.
14. Padhani AR. Dynamic contrast-enhanced MRI in clinical oncology: current status and future directions. *J Magn Reson Imaging*. 2002;16:407-422.
15. O'connor JPB, Rose CJ, Jackson A, et al. DCE-MRI biomarkers of tumour heterogeneity predict CRC liver metastasis shrinkage following bevacizumab and FOLFOX-6. *Br J Cancer*. 2011;105:139-145.
16. Yankeelov TE, Mankoff DA, Schwartz LH, et al. Quantitative imaging in cancer clinical trials. *Clin Cancer Res*. 2016;22:284-290.
17. Artzi M, Liberman G, Blumenthal DT, Bokstein F, Aizenstein O, Ben BD. Repeatability of dynamic contrast enhanced vp parameter in healthy subjects and patients with brain tumors. *J Neurooncol*. 2018;140:727-737.
18. Jaspan ON, Fleysher R, Lipton ML. Compressed sensing MRI: a review of the clinical literature. *Br J Radiol*. 2015;88:20150487.
19. Smith DS, Welch EB, Li X, et al. Quantitative effects of using compressed sensing in dynamic contrast enhanced MRI. *Phys Med Biol*. 2011;56:4933.
20. Feng L, Grimm R, Block KT, et al. Golden-angle radial sparse parallel MRI: combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI *Magn Reson Med*. 2014;72:707-717.
21. Rosenkrantz AB, Geppert C, Grimm R, et al. Dynamic contrast-enhanced MRI of the prostate with high spatiotemporal resolution using compressed sensing, parallel imaging, and continuous golden-angle radial sampling: preliminary experience. *J Magn Reson Imaging*. 2015;41:1365-1373.
22. Guo Y, Lingala SG, Zhu Y, Lebel RM, Nayak KS. Direct estimation of tracer-kinetic parameter maps from highly undersampled brain dynamic contrast enhanced MRI. *Magn Reson Med*. 2017;78:1566-1578.
23. Guo Yi, Lingala SG, Bliesener Y, Lebel RM, Zhu Y, Nayak KS. Joint arterial input function and tracer kinetic parameter estimation from undersampled dynamic contrast-enhanced MRI using a model consistency constraint. *Magn Reson Med*. 2018;79:2804-2815.
24. Dikaios N, Arridge S, Hamy V, Punwani S, Atkinson D. Direct parametric reconstruction from undersampled (k, t)-space data in dynamic contrast enhanced MRI. *Med Image Anal*. 2014;18:989-1001.
25. Bliesener Y, Acharya J, Nayak KS. Efficient DCE-MRI Parameter and Uncertainty Estimation Using a Neural Network. *IEEE Trans Med Imaging*. 2019;39:1712-1723.
26. Ulas C, Das D, Thrippleton MJ, et al. Convolutional neural networks for direct inference of Pharmacokinetic parameters: application to stroke dynamic contrast-enhanced MRI. *Front Neurol*. 2018;9:1147.
27. Ulas C, Tetteh G, Thrippleton MJ, et al. Direct estimation of Pharmacokinetic parameters from DCE-MRI using deep CNN with forward physical model loss. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2018:39-47.
28. Kettelkamp J, Lingala SG. Arterial input function and tracer kinetic model-driven network for rapid inference of kinetic maps in Dynamic Contrast-Enhanced MRI (AIF-TK-net). In: *IEEE 17th International Symposium on Biomedical Imaging (ISBI)*. IEEE; 2020:1450-1453.
29. Rastogi A, Yalavarthy PK. SpiNet: a deep neural network for Schatten p-norm regularized medical image reconstruction. *Med Phys*. 2021;48:2214-2229.
30. Aggarwal HK, Mani MP, Jacob M. MoDL: model-based deep learning architecture for inverse problems. *IEEE Trans Med Imaging*. 2018;38:394-405.
31. Zhang J, Ghanem B. ISTA-Net: interpretable optimization-inspired deep network for image compressive sensing. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE; 2018:1828-1837.

32. Rastogi A, Yalavarthy PK. Comparison of iterative parametric and indirect deep learning-based reconstruction methods in highly undersampled DCE-MR Imaging of the breast. *Med Phys*. 2020;47:4838-4861.
33. Venkatakrishnan SV, Bouman CA, Wohlberg B. Plug-and-play priors for model based reconstruction. In: *2013 IEEE Global Conference on Signal and Information Processing*. IEEE; 2013:945-948.
34. Yang Y, Sun J, Li H, Xu Z. Deep ADMM-Net for compressive sensing MRI. In: *Advances in neural information processing systems*. 2016:10-18.
35. Gan W, Shoushtari S, Hu Y, Liu J, An H, Kamilov US. Block Coordinate Plug-and-Play Methods for Blind Inverse Problems *arXiv preprint arXiv:2305.12672 (accepted in NeurIPS 2023)*. 2023.
36. Ulyanov D, Vedaldi A, Lempitsky V. Deep image prior. *arXiv preprint arXiv:1711.10925*. 2017;doi:10.48550/arXiv.1711.10925
37. Rahaman N, Baratin A, Arpit D, et al. On the spectral bias of neural networks. In: *International Conference on Machine Learning*. PMLR; 2019:5301-5310.
38. Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. *Phys D: Nonlinear Phenom*. 1992;60:259-268.
39. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13:600-612.
40. Chan SH, Wang X, Elgendy OA. Plug-and-play ADMM for image restoration: fixed-point convergence and applications. *IEEE Trans Comput Imaging*. 2016;3:84-98.
41. Liu DC, Nocedal J. On the limited memory BFGS method for large scale optimization. *Math Program*. 1989;45:503-528.
42. Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn Reson Med*. 2007;58:1182-1195.
43. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer; 2015:234-241.
44. Clark K, Vendt B, Smith K, et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. 2013;26:1045-1057.
45. Huang W, Li X, Chen Y, et al. Variations of dynamic contrast-enhanced magnetic resonance imaging in evaluation of breast cancer therapy response: a multicenter data analysis challenge. *Transl Oncol*. 2014;7:153-166.
46. Barboriak D. Data from RIDER_NEURO_MRI. *The Cancer Imaging Archive*. 2015;577.
47. Fedorov A, Tempany C, Mulkern R, Fennessy F. Data from QIN PROSTATE. *Cancer Imaging Arch*. 2016;doi:10.7937/K9/TCIA.2016.fADs26kG
48. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. 2014;doi:10.48550/arXiv.1412.6980
49. Huang J, Zhang S, Metaxas D. Efficient MR image reconstruction for compressed MR imaging. *Med Image Anal*. 2011;15(5):670-679.doi:10.1016/j.media.2011.06.001
50. Okawa S, Hoshi Y. A review of image reconstruction algorithms for diffuse optical tomography. *Appl Sci*. 2023;13:5016.
51. Kupis S. *Methods for the Electrical Impedance Tomography Inverse Problem: Deep Learning and Regularization with Wavelets*. PhD thesis. Clemson University; 2021.

How to cite this article: Rastogi A, Yalavarthy PK. Greybox: A hybrid algorithm for direct estimation of tracer kinetic parameters from undersampled DCE-MRI data. *Med Phys*. 2024;1-21. https://doi.org/10.1002/mp.16935