RPE | Radiation Physics & Engineering
A Peer-Reviewed Journal
Published by K.N. Toosi University of Technology
http://rpe.kntu.ac.ir
Research Article

Radiation Physics and Engineering 2026; ?(?):?–?

PKest-LSnet: A robust LSTM-Based deep learning framework for accurate pharmacokinetic parameter estimation in DCE-MRI with clinical validation

Azimeh NV Dehkordi[a,*], Hossein Malekmohammadi[b], Sedigheh Sina[b,c]

[a] *Department of Physics, Na. C, Islamic Azad University, Najafabad, Iran*
[b] *Mechanical Engineering Department, Shiraz University, Shiraz, Iran*
[c] *Radiation Research Center, Shiraz University, Shiraz, Iran*

HIGHLIGHTS

- The model maintained stability across variable acquisition protocols.
- Excelling in tumor characterization (DSC=77%) leaky vasculature classification challenges (DSC=36%).
- This work bridges the gap between simulated training and clinical deployment.

ABSTRACT

Dynamic contrast-enhanced MRI (DCE-MRI) plays a pivotal role in quantifying tissue hemodynamics, yet accurate pharmacokinetic (PK) parameter estimation remains challenging. We present PKest-LSnet, a novel Long Short-Term Memory (LSTM)-based deep learning framework for robust PK analysis. The network was trained on comprehensive simulated data generated using physiologically plausible ranges of PK parameters (parameetrs of Tofts equation: v_p, K^{trans}, k_{ep}) and a population-based arterial input function (AIF), enabling precise ground-truth validation. Rigorous testing on clinical data from 19 glioblastoma patients demonstrated strong agreement with reference method maximum likelihood estimation (MPE <12% for all parameters), with 99.42% accuracy in classifying tumor signals. The model maintained stability across variable acquisition protocols (time intervals: 2-6 sec; SNR: 5-100), proving its adaptability to real-world clinical variability. While excelling in tumor characterization (DSC=77.27%), leaky vasculature classification challenges (DSC=36.76%) revealed opportunities for architectural refinements. PKest-LSnet eliminates dependency on initial parameter estimates, reduces computational time by orders of magnitude compared to conventional methods, and offers a robustness method to the tested variations in temporal sampling intervals and SNRs-critical for multicenter studies. This work bridges the gap between simulated training and clinical deployment, providing a validated tool for precision DCE-MRI analysis with direct applications in oncology and therapeutic monitoring. Future directions include hybrid architectures for improved intermediate-tissue classification and patient-specific AIF integration.

KEYWORDS

DCE-MRI
Pharmacokinetic modeling
LSTM networks
Glioblastoma
Deep learning

HISTORY

Received:
Revised:
Accepted:
Published:

1 Introduction

Pharmacokinetic (PK) parameters serve as critical biomarkers for diagnosing diseases and monitoring cancer treatment, offering insights into tissue microvasculature and treatment response (Li et al., 2015; Bagher-Ebadian et al., 2011; Rygh et al., 2014; Gravina et al., 2019). These parameters provide prognostic value for various conditions including stroke, tumor progression, and therapeutic resistance, making their accurate quantification essential in clinical practice (Reavey-Cantwell et al., 2001; Choi et al., 2015; Metzler, 1986).

Dynamic Contrast-Enhanced MRI (DCE-MRI) has emerged as a key non-invasive technique for PK parameter measurement, tracking contrast agent behavior through temporal changes in T_1 relaxation times (Tofts et al., 1999; Tofts and Kermode, 1991; Lonser et al., 2019). Current estimation methods fall into two categories: model-free

*Corresponding author: nourizadeh@iau.ac.ir

https://doi.org/

approaches that analyze signal curve characteristics, and model-based techniques that fit mathematical models to the data (Tofts and Kermode, 1991; Urien and Lokiec, 2004; Acharjee et al., 2014; Dansirikul et al., 2005; Hsu et al., 2013). Among these, Tofts model (Tofts et al., 1999; Tofts and Kermode, 1991; Lonser et al., 2019) has gained particular prominence in brain studies for its effective characterization of vascular permeability (K^{trans}), plasma volume (v_p), and extracellular space (v_e)- all clinically significant parameters.

Traditional estimation employs least squares techniques (Levenberg-Marquardt, Gaussian-Newton, etc.) (Bagher-Ebadian et al., 2016; Gill et al., 2015; Wang et al., 2012; Yu and Cao, 2017; Brocks and Hamdy, 2020), but these methods face several limitations: computational intensity, sensitivity to initial values, and vulnerability to noise and acquisition parameters (Metzler, 1986; De Naeyer et al., 2011; Schabel and Parker, 2008; Yan et al., 2014). These challenges have driven the adoption of deep learning approaches which offer distinct advantages (Klepaczko et al., 2020; Dehkordi et al., 2017, 2021).

Neural networks excel in analyzing DCE-MRI data by detecting complex patterns beyond conventional methods' reach (Jung et al., 2022; Oh et al., 2024; Ottens et al., 2022; Simeth and Cao, 2020; Xue et al., 2019). Unlike traditional approaches, deep learning eliminates manual parameter tuning, achieving optimized solutions through training. This has expanded their use in medical imaging for improved diagnosis, prognosis, and biomarker discovery. Deep learning marks a major leap in PK parameter estimation, overcoming prior limitations and unlocking new clinical applications. These data-driven methods enhance computational efficiency and robustness across varying imaging conditions. As the field advances, deep learning is set to revolutionize quantitative DCE-MRI analysis, enabling more precise, personalized tissue assessment for better patient outcomes (Ulas et al., 2019, 2018; Zou et al., 2020).

Ulas et al. (Ulas et al., 2019) developed a 2D Convolutional Neural Network (CNN) for direct PK parameter estimation from DCE-MRI time curves, eliminating conventional processing steps. While innovative, this arterial input function (AIF)-independent approach had limitations since PK accuracy fundamentally depends on AIF characterization (Ulas et al., 2019). Ziayee et al. (Ziayee et al., 2018) subsequently enhanced this by incorporating AIF as an additional CNN input, improving K^{trans} and Kep estimation (though v_p showed minimal gains). Both CNN approaches faced inherent constraints in modeling long-term hemodynamic temporal relationships. Addressing this, Zou et al. (Zou et al., 2020) implemented an Long Short-Term Memory (LSTM) network processing both concentration-time curves and patient-specific AIF, achieving superior accuracy and computational efficiency compared to CNN methods by better capturing temporal dependencies in the sequential DCE-MRI data.

Inspired by the successes of LSTM networks and the limitations of CNN-based approaches, this study proposes an LSTM-based network designed to extract long and short time-dependent features from DCE-MRI time signals. Our LSTM-based innovation simultaneously performs tissue pathology classification and PK parameter estimation using population-averaged AIF, addressing clinical variability through data augmentation with different SNRs and sampling intervals. The network was optimized via hyperparameter tuning and validated on both simulated and clinical GBM data, demonstrating robust performance across acquisition protocols. Key advantages include: 1) combined model selection/parameter estimation, 2) handling of protocol variations, and 3) faster computation versus traditional methods. This represents a significant advance over both conventional fitting techniques and earlier deep learning implementations

2 Material and Methods

2.1 Data Preparation

Preparing the training dataset is a crucial step in developing deep learning-based networks. The accuracy of LSTM-based PK parameter estimation depends on the quality and representativeness of the training data (Donahue et al., 2015). The dataset must be sufficiently large and diverse to enhance the model's generalization capability. Using simulated data offers advantages such as precise ground truth labels, which can be costly or difficult to obtain from real clinical datasets. This approach enables more accurate model evaluation and validation while providing full control over data generation, improving reproducibility and allowing the study of various performance-influencing factors.

In this study, two datasets were used: a synthetic dataset for training and evaluating the LSTM-based PKest-LSnet and a real-world dataset from glioblastoma (GBM) patients to test the model's performance on clinical data.

2.1.1 Synthetic Data

The study employed Tofts' equation (Eq. (1)) to simulate DCE-MRI data, modeling contrast agent concentration (C_t) as a function of arterial input (C_p/AIF) and three pharmacokinetic parameters: plasma volume (v_p), forward transfer constant (K^{trans}), and backward transfer constant (kep). The Extravascular Extracellular Space (EES), denoted as (ve), can be calculated from the ratio K^{trans}/k_{ep} (Tofts et al., 1999):

$$C_t(t) = v_p C_p(t) + K^{trans} \int C_p(\tau) e^{-k_{ep}(t-\tau)} \mathrm{d}\tau \qquad (1)$$

Three nested models were derived based on tissue physiology: Model 1 (normal tissue, Eq. (2)) using only v_p; Model 2 (leaky vasculature, Eq. (3)) with v_p and K^{trans}; and Model 3 (complete exchange, Eq. (4)) incorporating all three parameters. A population-averaged AIF was created from expert-selected patient-specific AIFs across 19 GBM cases.

$$\textbf{Model 1:} \quad C_t(t) = v_p C_p(t) \qquad (2)$$

$$\textbf{Model 2:} \quad C_t(t) = v_p C_p(t) + K^{trans} \int C_p(\tau) \mathrm{d}\tau \qquad (3)$$

Table 1: Considered range for each PK parameter in simulation.

Model	Range	Number of Generated Signals
Model #1	$V_p = [0.00075{:}1.6]$ $K^{trans} = 0$ $K_{ep} = 0$	4320
Model #2	$V_p = [0.00075{:}1.6]$ $K^{trans} = [0.03{:}2]$ $K_{ep} = 0$	285120
Model #3	$V_p = [0.00075{:}1.6]$ $K^{trans} = [0.015{:}2]$ $K_{ep} = [0.015{:}2]$	574560

$$\textbf{Model 3:}\quad C_t(t) = v_p C_p(t) + K^{trans}\int C_p(\tau)e^{-k_{ep}(t-\tau)}\mathrm{d}\tau \tag{4}$$

To create a synthetic dataset that accurately represents the measured patient data, a wide and physiologically relevant range of PK parameters (as detailed in Table 1) was employed to simulate the contrast concentration curves based on Tofts equation.

Consequently, 4320 signals were generated for Model 1, 285120 signals for Model 2, and 574560 signals for Model 3. These datasets were utilized as synthetic data for training the PKest-LSnet. The use of simulated data during the training process allows for precise calculations of the mean percentage error for each estimated PK parameter, as the ground truth values are well-defined and the simulated signals are generated using known PK parameters.

2.1.2 Patient Data

The PKest-LSnet was validated using clinical DCE-MRI data from 19 GBM patients obtained from public dataset at The Cancer Imaging Archive (TCIA) (Clark et al., 2013). This real-world evaluation tested the LSTM network's ability to process actual clinical inputs and generate reliable outputs. All patients underwent 1.5T MRI scans with the following acquisition parameters: For T_1 mapping, 3D FLASH images used multiple flip angles (5-30°), $TR/TE = 4.43/2.1$ ms, and two signal averages. Dynamic imaging during contrast injection (0.1 mmol.kg^{-1} Magnevist at 3 cc.s^{-1}) employed a 25° flip angle, $TR/TE = 3.8/1.8$ ms, $1 \times 1 \times 5$ mm voxels, and 16 slices acquired every 4.8 seconds starting 24 seconds post-scan initiation. This comprehensive clinical validation ensured the model's robustness with real patient data under standard imaging conditions.

- Preprocessing of Clinical Data

The analysis involved preprocessing patient DCE-MRI data by first cropping images to brain boundaries and generating time-intensity signals for each voxel. Native T_1 (T_{10}) maps were created, and longitudinal relaxation rate changes (ΔR_1) were calculated using Eq. (5) (Tofts and Kermode, 1991).

$$\Delta R_1 = \frac{1}{T_1} - \frac{1}{T_{10}} \tag{5}$$

Tissue contrast agent concentration (C_t) was derived by multiplying ΔR_1 by $(1 - H_{ct})$, with hematocrit (H_{ct}) fixed at 0.45. These concentration-time curves served as input to the PKest-LSnet network. Voxels with mean signal $< 1^{-3}$ were excluded as non-enhancing regions. This preprocessing pipeline ensured meaningful pharmacokinetic analysis while filtering out unenhanced brain tissue.

- Ground Truth Maps Preparation for Clinical Data

To estimate the percentage error of each estimated parameter for clinical data, ground truth maps of PK parameters for each voxel is required. Therefore, the clinical GBM patient data were accompanied by maps of PK parameters estimated using the Maximum Likelihood Estimation (MLE) technique (Bagher-Ebadian et al., 2016), which serves as the reference method. MLE is recognized for its accuracy, robustness, and low bias (Bagher-Ebadian et al., 2016). In this study, an MLE-based algorithm was employed to estimate the PK parameters for all voxels within the brain. However this point is good to be mentioned that for the clinical data, MPE represents the relative deviation of the proposed estimates from the MLE-based reference estimates and should therefore not be interpreted as an absolute estimation error.

2.1.3 Augmentation

After training the network with the base data and specifying the hyperparameters, it was necessary to enhance the network's robustness against noise, to ensure its applicability across a wide range of patient data acquired from various scan centers (NV and Koohestani, 2019). To achieve this, augmentation techniques were employed. White Gaussian noise at various levels (SNR: 5, 10, 15, 20, 25, 50 and 100.) was added to the pure data to generate several additional noisy datasets. Each noisy dataset, along with the pure data, was utilized to train the network. This approach aimed to investigate how network performance varied in the presence of noisy data alongside the base data.

2.2 PKest_LSnet Structure

LSTM, a recurrent neural network variant, effectively models complex contrast agent (CA) concentration-time relationships for precise PK parameter estimation. Trained on CA concentration-time profiles, it captures long-term dependencies in sequential data, enabling accurate predictions even with irregular sampling and time lags (Zhao et al., 2020). Our proposed PKest_LSnet method frames PK estimation as a mapping from CA concentration curves (with AIF) to PK parameters, using the extended Tofts model. The synthetic data generation process is shown in Fig. 1, while Fig. 2 illustrates PKest_LSnet's architecture and evaluation framework. This approach demonstrates LSTM's capability for robust pharmacokinetic analysis in dynamic contrast-enhanced imaging.

As depicted in this figure, the developed network comprises four LSTM layers, each containing a number of LSTM cells equal to the length of the input signal.

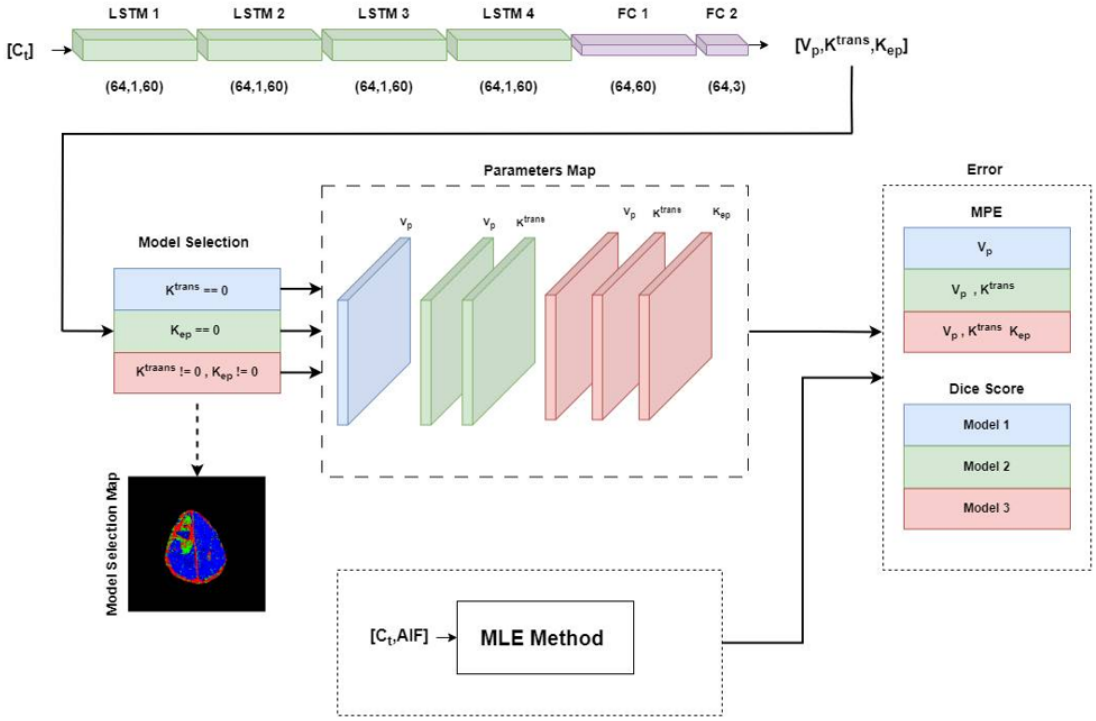

Figure 1: The generation process of the synethic data.

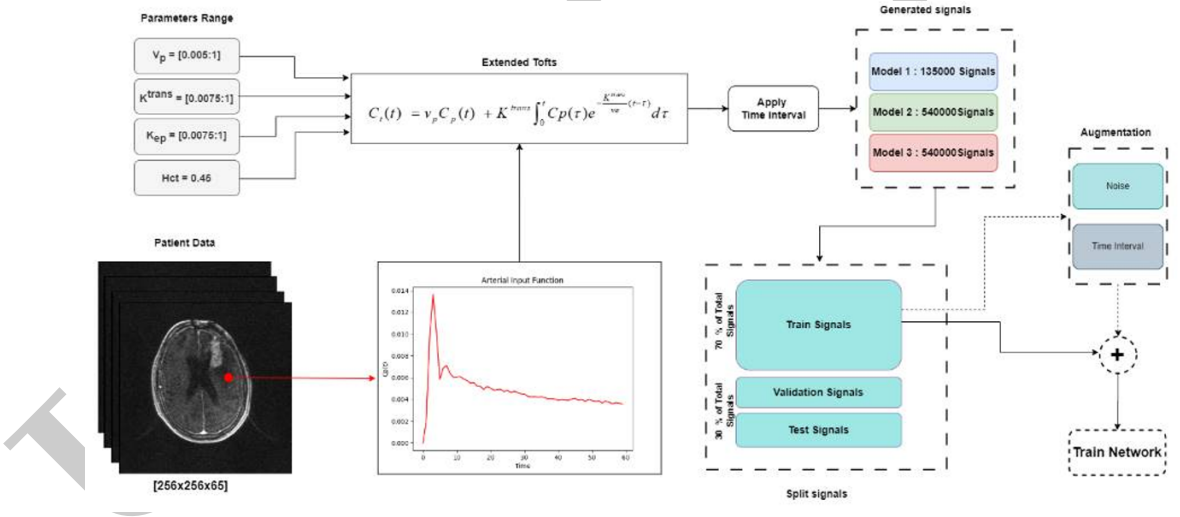

Figure 2: the structure of the developed PKest-LSnet and its evaluation process.

The CA time-concentration signals, consisting of 60 time points, served as the network's input. The input signals were arranged as a 3D matrix with dimensions (Batch size, 1, number of cells in each layer). Input signals are processed within each LSTM cell, and the hidden state (h_t) is extracted using the following formulas and transferred to the corresponding cell in the next layer (Eqs (6) to (10)) (Xue et al., 2019):

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{6}$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{7}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{8}$$

$$c_t = f_t \times c_{t-1} + i_t \times \tanh\left(W_c x_t + U_c h_{t-1} + b_c\right) \tag{9}$$

$$h_t = o_t \times \tanh(c_t) \tag{10}$$

The LSTM architecture processes temporal data through sequential operations, where ht-1 represents the previous hidden state and Xt is the timepoint-specific input. The network employs sigmoid activation (σ), input weights (W), hidden state weights (U), and biases (b) to compute cell outputs (c_t). The final LSTM layer output passes through a 60-node fully connected layer (matching input signal length) with ReLU activation to generate the extended Tofts model PK parameters. The network first estimates voxel-wise PK parameters, then determines the appropriate tissue model (Model 1/2/3) based on non-zero parameter values, enabling simultaneous parameter estimation and model selection.

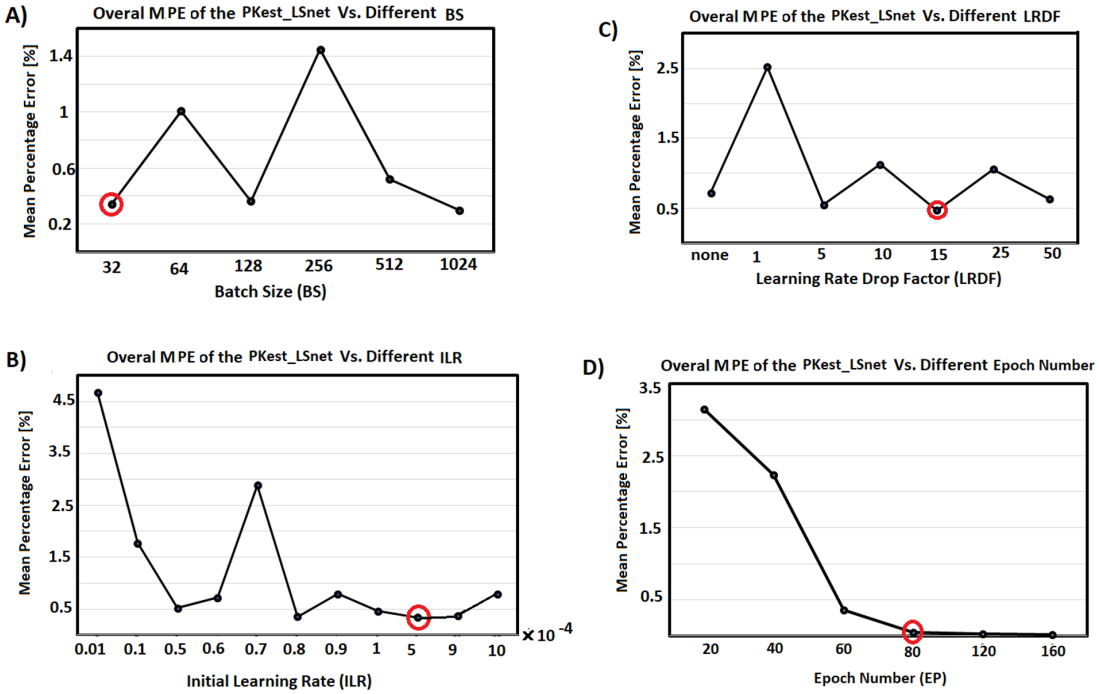

Figure 3: The Impact of hyperparameters optimization on PKest-LSnet Estimation Accuracy.

2.3 Training and Optimization of the PKest-LSnet

Hyperparameter optimization was critical for developing the LSTM network, with extensive tuning performed on: (1) architectural parameters (number of LSTM layers/units), (2) training parameters (initial learning rate, batch size, epochs), and (3) regularization factors (dropout rate, learning rate decay schedule). The random search method identified optimal values balancing model capacity (to capture complex temporal patterns) against computational efficiency and overfitting risks. Key trade-offs included: increasing LSTM units improved temporal modeling but raised computational demands, while aggressive learning rates accelerated convergence but risked instability.

The final optimized network - using 70% synthetic data for training, 15% for validation, and 15% for testing - used for PK parameter prediction for GBM patients by simultaneously optimizing all interdependent hyperparameters.

As described above, the base synthetic concentrationtime curves were generated using the parameter ranges listed in Table 1. To ensure independence among the training, validation, and test datasets and to prevent information leakage, the base synthetic signals were first randomly divided into three mutually exclusive subsets: 70% for training, 15% for validation, and 15% for final testing. This partitioning was performed before any data augmentation or temporal resampling. Subsequently, noise augmentation was applied independently within each subset, whereas signals with different temporal sampling intervals were generated exclusively for the test set to evaluate the model's robustness to variations in temporal resolution. Accordingly, all noisy versions derived from a given base signal remained within the same subset as the original signal, ensuring that no version of a test signal was used during training or validation. The augmented noisy data were therefore included in the training, validation, and testing processes, while the temporally resampled signals with different sampling intervals were used exclusively for testing.

2.4 Loss Function

For the loss function, we employed the Mean Absolute Error (MAE) (refer to Eq. (11)) and Mean Squared Error (MSE) (refer to Eq. (12)) to assess the network's performance. During each training session, we calculated the weighted average absolute error and mean suared error between the gold standard maps of PK parameters and the estimated maps produced by the network. Subsequently, based on the optimizer used, the network's weights were updated to minimize the average absolute error (Kayastha et al., 2023).

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{11}$$

$$MSE = \frac{\sum_{i=1}^{n} (y_i - x_i)^2}{n} \tag{12}$$

2.5 Evaluating Metrics

The networks weights are adjusted based on the difference between true parameter values and its output. Alongside MSE, Mean Absolute Percentage Error (MAPE) was computed for training and validation data. For real patient data evaluation, the Dice Similarity Coefficient (DSC) was also added to compare model selection maps from LSTM outputs and MLE-LLR framework.

The sampling interval of CA concentration-time between consecutive imaging acquisitions-varies by center and sequence. To test robustness, the network was evaluated using simulated signals with different sampling intervals (TIs: 2-6 seconds) and noise levels (SNRs: 5-100). Baseline simulations used a 4.8-second TI, matching clinical data. Noise augmentation's impact on classification accuracy and PK parameter estimation was assessed by testing networks trained with and without added noise. All algorithms ran on TensorFlow using an Intel Xeon CPU, 16GB RAM, and an NVIDIA Tesla P100 GPU.

3 Results

3.1 PKest-LSnet Optimization Results

The network's hyperparameters were optimized via random search to maximize accuracy and convergence speed. Key parameters were systematically evaluated across wide ranges, with performance assessed through all three models' PK estimation accuracy. This comprehensive optimization process identified optimal values that balanced computational efficiency with precise parameter estimation across different tissue types (Models 1-3). The method ensured robust network performance while maintaining clinical relevance for diverse physiological conditions.

Figures 3-A to 3-D show how hyperparameters affect PK parameter estimation accuracy (MPE). Figure 3-A reveals MPE decreases from ∼0.3% to ∼0.2% as batch size increases from 32 to 1024, with larger batches stabilizing training through reduced gradient variance. Despite 1024 achieving lowest MPE, practical computational constraints justified selecting BS=32 ($MPE \sim 0.3\%$) as optimal. Figure 3B demonstrates learning rate effects, with $ILR = 5e-4$ yielding minimum MPE (∼1.5%). Higher rates ($> 5e-4$) caused instability ($MPE > 4.5\%$), while lower rates slowed convergence without accuracy gains. These findings guided optimal hyperparameter selection for network training, balancing computational efficiency with estimation accuracy.

Figures 3-C evaluates learning rate drop factor (LRDF) and training epochs. Testing LRDF values (0.1-0.9) at various intervals showed 0.4 (every 15 epochs) minimized MPE (∼0.5%). Larger factors (e.g., 0.9) or no adjustment degraded performance. Figure 3-D reveals training stabilizes after 80 epochs ($MPE \sim 1.5\%$), with no significant improvement beyond. To optimize memory usage while maintaining accuracy, 80 epochs was selected as optimal. These results demonstrate the importance of carefully tuned learning rate decay and sufficient (but not excessive) training duration for achieving stable, accurate parameter estimation.

Thus, the final configuration used batch size 32, ADAM optimizer, an initial learning rate of $5e-4$ reduced by 0.4 every 15 epochs, and 80 training epochs with 1000 iterations each. This balanced approach optimized both computational efficiency and estimation accuracy.

Table 2: Correct model selection fraction and MPE of the estimated parameters for each model.

Model	MPE of Estimated Models Parameters		Correct Model Classification Fraction
Model #1	MPE of v_p	0.03%	100%
Model #2	MPE of v_p	0.3%	75.8%
	MPE of K^{trans}	0.2%	
Model #3	MPE of v_p	0.08%	99.4%
	MPE of K^{trans}	0.87%	
	MPE of k_{ep}	1.1%	

Table 3: Confusion matrix of true recognization of Models for each pixel.

		Predicted		
		Model 1	Model 2	Model 3
True	Model 1	100.0	0.0	0.0
	Model 2	0.15	75.78	24.06
	Model 3	0.55	0.029	99.42

3.2 Evaluation Results of the PKest-LSnet

3.2.1 Network Evaluation with Syntethic Data

PKest-LSnet showed excellent performance on unseen simulated data (15% test set), accurately classifying signal models and estimating parameters. As Table 2 demonstrates, it achieved remarkable precision for both Model 1 (normal tissue, v_p MPE=0.03%) and Model 3 (tumor regions). For clinically critical Model 3, all parameters (v_p, K^{trans}, kep) were estimated with $< 1.1\%$ MPE. The network's exceptional accuracy in tumor-related Model 3 classification and parameter estimation is particularly valuable, as this model represents the primary focus in most clinical and research applications.

Table 3's confusion matrix reveals PKest-LSnet's classification performance for tissue types using extended Tofts model parameters. The network achieved perfect accuracy (100%) for Model 1 (normal tissue) and near-perfect results (99.42%) for Model 3 (tumor). However, Model 2 (reversible leakage) showed reduced accuracy (75.78%), with 24.06% misclassified as Model 3, reflecting the challenge in distinguishing these similar leakage patterns. While demonstrating excellent tumor detection and normal tissue identification capabilities, these results indicate the need for improved differentiation between reversible (Model 2) and irreversible (Model 3) exchange mechanisms, potentially through enhanced feature extraction or expanded training data for intermediate leakage cases.

3.2.2 Network Evaluation with Real Patient Data

The PK parameter estimation Table 4 compares LSTM predictions against the gold-standard MLE technique across 19 GBM patients. For Model 1, v_p estimates showed excellent agreement with MPE of 6.83% and low variability. Model 2 maintained consistent v_p estimation (MPE=8.06%) but showed higher error in K^{trans}

Table 4: Averaged PK parameters estimated by LSTM and MLE (as reference method) for all recruited GBM patients in this study.

PK Estimated values by LSTM and their MPE related to the MLE technique						
	Model 1	Model 2		Model 3		
Technique	v_p[%] Mean±SD	v_p[%] Mean±SD	K^{trans}[min^{-1}] Mean±SD	v_p[%] Mean±SD	K^{trans}[min^{-1}] Mean±SD	K_{ep}[min^{-1}] Mean±SD
LSTM	0.0109±0.0081	0.0228±0.0149	0.0465±0.0153	0.0356±0.0266	0.0558±0.0152	0.353±0.134
MLE	0.0117±0.0112	0.0248±0.0174	0.0510±0.0223	0.0342±0.0154	0.0612±0.0034	0.317±0.187
MPE [%]	6.83%	8.06%	8.8%	4.09%	8.7%	11.3%

(MPE=8.8%), potentially due to noise sensitivity in low-permeability tissues. Model 3 demonstrated the best performance for v_p (MPE=4.09%) while kep had moderate error (MPE=11.3%), reflecting inherent challenges in estimating backward transfer rates. Overall, the LSTM achieved clinical-grade accuracy with all parameters below 12% MPE.

In this study, the MLE-LLR framework was used as a statistically established reference rather than an absolute gold standard. Thus, in clinical data, the reported MPE and DSC values reflect the agreement of the proposed method with the reference framework rather than validation against a known biological ground truth. MLE may also be influenced by noise, model misspecification, parameter identifiability, and optimization-related factors, which should be considered when interpreting the clinical comparisons.

DSC scores evaluating spatial agreement between LSTM and MLE maps revealed Model 1 had exceptional performance (average DSC=93.19%), crucial for excluding healthy tissue in diagnostics. Model 2 had the lowest DSC (36.76%), highlighting difficulties in delineating early-stage leakage that may require more training data. Model 3 showed reliable performance (DSC=77.27%) for tumor assessment, with variations (e.g., Patient 3: 63.54%) likely reflecting heterogeneous enhancement patterns. The obtained results from the real patient data is matched with the obtained results from the synthetic data, please see the Table 5 for more detailes.

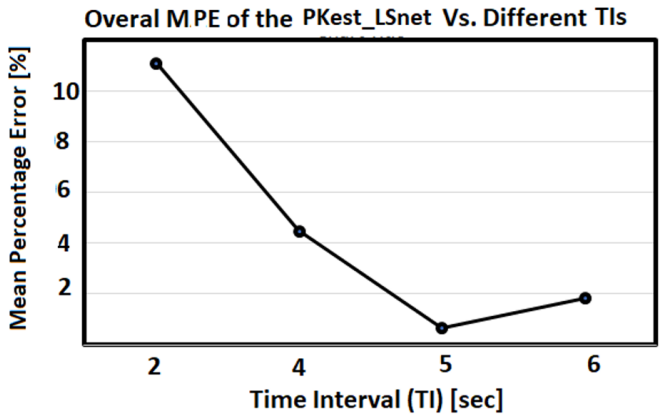

Figure 4: Overal MPE of the trained PKest_Lsnet tested against the simulated signals with various time Intervals (TIs).

Table 5: Dice Similarity Coefficient between the Model Selection By Trained PKest$_L$*SnetandMLEbasedmethods*.

	DSC score		
Patient	Model#1	Model#2	Model#3
1	97.00	39.70	84.40
2	95.60	30.40	75.00
3	89.13	32.11	63.54
4	96.90	32.90	84.50
5	96.30	38.20	74.70
6	97.70	49.00	66.10
7	94.80	31.90	76.10
8	91.60	35.50	86.20
9	96.40	47.70	85.10
10	94.90	38.50	67.50
11	97.30	29.56	79.3
12	92.80	38.70	65.75
13	96.60	32.00	90.80
14	95.80	46.80	78.30
15	93.40	42.60	66.70
16	88.10	33.20	87.70
17	95.10	37.80	85.10
18	75.90	32.30	81.70
19	85.40	29.70	69.80
Average	93.19	36.76	77.27

The lower performance of Model 2 may be attributed to its limited spatial representation and its similarity to Model 3. Model-2 regions are typically small and spatially scattered around the tumor, resulting in a limited number of available signals for learning their distinctive characteristics. Moreover, Models 2 and 3 may exhibit similar temporal enhancement patterns, particularly during the early enhancement phase, making their discrimination more challenging. This is consistent with the higher rate of Model-2 signals misclassified as Model 3 in the synthetic dataset and may also contribute to the lower DSC observed for Model 2 in the clinical data.

3.3 Test of Network Robustness to Variations in Temporal Intervals and SNRs

The robustness of PKest-LSnet to variations in noise levels and temporal sampling intervals (TIs) was evaluated using the augmented synthetic test data. Figure 4 presents the overall MPE obtained at different TIs, demonstrating the network's performance under varying temporal sampling conditions. The lowest MPE ($\sim$ 0.5%) was observed at the baseline TI of 5 s, which corresponds to the temporal sampling condition used for the base data. The perfor-

mance remained relatively stable at moderately different TIs (4-6 s), with MPE values below 6%, indicating good generalization to the tested variations in temporal sampling intervals. However, the MPE increased to approximately 11% at the more substantial deviation of TI = 2 s, indicating reduced performance under more extreme temporal sampling conditions. Overall, these results demonstrate robustness to the tested range of temporal sampling intervals while also indicating that performance may deteriorate when the temporal resolution substantially deviates from the baseline condition. These findings should be interpreted within the scope of the simulated conditions evaluated in this study and do not, by themselves, establish protocol-independent or multicenter robustness.

Figures 5-A to 5-C and 4-A to 4-C demonstrate the robustness of PKest-LSnet for PK parameter estimation and model selection under varying SNR conditions. The mean percentage error (MPE) analysis (Fig. 4-A) compares two training strategies: (1) training with both noise-free and noise-augmented data and (2) training exclusively with noise-free data. The results show that, without noise augmentation, the estimation error increases as the SNR decreases. In contrast, training with noise-augmented data substantially reduces the sensitivity of the network to noise and maintains more stable estimation performance across the tested SNR range.

Importantly, the test dataset included noise conditions that were not represented during training. In particular, an intermediate SNR condition (SNR=17) was excluded from the training set and was used only during testing, providing an additional assessment of the network's ability to generalize to an unseen noise level. Thus, the observed performance was not solely evaluated under noise conditions directly encountered during training.

The model selection accuracy shown in Figs. 4-B to 4-C follows a similar trend. Noise augmentation improves model classification, particularly under lower-SNR conditions, whereas the network trained exclusively on noise-free data shows a greater degradation in classification performance as the noise level increases. A few unseen SNRs during training process were also used in testing the model selection accuracy. These findings indicate that exposure to a range of noise conditions during training improves the stability of both PK parameter estimation and model selection in the presence of noise.

Overall, the results demonstrate that incorporating noise augmentation enhances the robustness of PKest-LSnet to the tested SNR conditions and improves its ability to generalize to noise levels not explicitly represented during training. However, these findings were obtained using simulated data and should be interpreted as evidence of robustness to the tested noise conditions rather than as evidence of protocol-independent or multicenter performance. Further validation using clinical datasets acquired with different scanners, imaging protocols, and institutions would be required to establish the broader generalizability of the proposed framework.

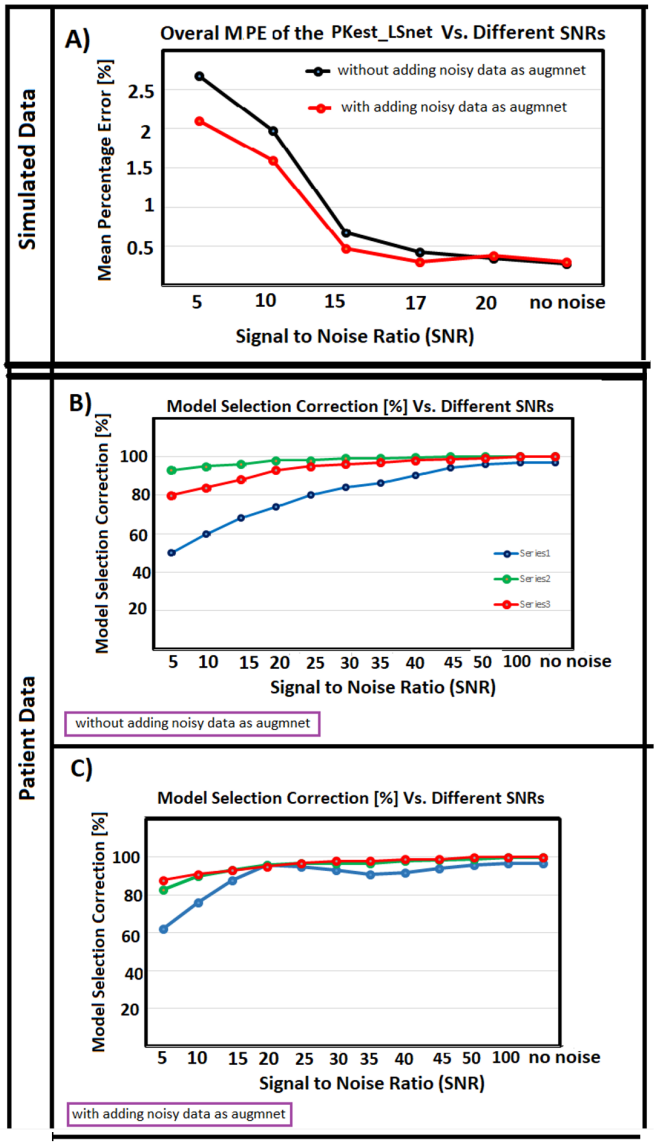

Figure 5: This figure demonestrates the robustness of the trained network (PKest_LSnet) for PK estimation over different SNRs.

4 Discussion

PKest-LSnet demonstrates its potential as a robust DCE-MRI analysis tool, achieving accuracy comparable to traditional methods while improving computational efficiency and addressing key quantitative imaging challenges (Dansirikul et al., 2005; Gill et al., 2015; Dehkordi et al., 2017; Hỳvlová et al., 2024). The network's performance may be attributed to its LSTM architecture, which captures temporal patterns in contrast-agent kinetics and provides an effective framework for PK estimation (Tofts et al., 1999; Ulas et al., 2018). Unlike conventional approaches that may require appropriate initial parameter estimates, PKest-LSnet estimates PK parameters without such initialization while substantially reducing processing time.

The use of comprehensive simulated data provided three key advantages: (1) precisely known ground-truth labels for quantitative assessment of estimation errors, (2) controlled variation of PK parameter ranges and

noise characteristics, and (3) systematic evaluation under diverse physiological and acquisition-related conditions. The simulations incorporated physiologically plausible PK parameters and a population-averaged AIF derived from 19 GBM patients, providing clinically relevant input characteristics for network training.

Evaluation using data from 19 glioblastoma patients demonstrated strong performance, with a mean error of $< 12\%$ relative to the reference MLE estimates. The network achieved 99.42% classification accuracy for Model 3, whereas Model 2 showed lower performance, with a DSC of 36.76%. This lower performance may be related to the limited spatial representation of Model 2 and its temporal similarity to Model 3, which can make their discrimination more challenging. Future improvements may therefore benefit from larger datasets and architectures that incorporate both spatial and temporal information.

Automatic classification of tissue into three pharmacokinetic models may provide several potential advantages, including standardized tissue characterization, improved inter-reader consistency, quantitative PK parameters as potential biomarkers of treatment response, and near real-time analysis for clinical decision support.

The network demonstrated stable performance across the tested range of temporal sampling intervals and SNR conditions, indicating robustness to the simulated variations in temporal resolution and noise level. The inclusion of different SNRs and temporal sampling intervals was intended to represent important sources of acquisition-related variability that can affect the appearance and temporal characteristics of DCE-MRI concentration-time curves. In particular, the robustness analysis included noise conditions not directly represented during training, providing an additional assessment of generalization to unseen noise levels.

Nevertheless, these robustness results should be interpreted within the scope of the simulated conditions evaluated in this study. Although the simulations incorporated several important sources of variability, they cannot fully reproduce the complexity of real clinical DCE-MRI data. The clinical evaluation was limited to 19 patients acquired using a single imaging protocol; therefore, the present results do not establish protocol-independent or multicenter generalizability. Further validation using independent datasets from different scanners, imaging protocols, and clinical centers is required. Although simulations provide precisely known ground-truth parameters, differences between simulated and real tissue kinetics may remain. Patient-specific AIFs may further improve clinical parameter estimation beyond the population-averaged AIF used in this study.

5 Conclusions

PKest-LSnet provides an accurate and computationally efficient framework for DCE-MRI pharmacokinetic analysis, with its performance demonstrated through both synthetic and clinical evaluations. The LSTM-based architecture effectively captures temporal contrast-agent kinetics, enabling rapid PK parameter estimation and automatic pharmacokinetic model classification. Robustness analyses demonstrated stable performance across the tested variations in SNR and temporal sampling intervals, including selected conditions not directly represented during training. These findings support the potential of PKest-LSnet as a practical tool for quantitative DCE-MRI analysis and motivate further evaluation using larger and more diverse clinical datasets. Future validation across different scanners, acquisition protocols, and clinical centers is warranted to establish the broader generalizability and clinical applicability of the proposed framework.

Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work.

Funding

The authors declare that no funds, grants, or other financial support were received during the preparation of this manuscript.

References

Acharjee, M. K., Kumar, U., Bachar, S. C., et al. (2014). Estimation of pharmacokinetic parameters using nonlinear fixed effects one compartment open model. *Journal of Applied Pharmaceutical Science*, 4(12):056–059.

Bagher-Ebadian, H., Dehkordi, A., and Ewing, J. (2016). SU-F-I-26: Maximum likelihood and nested model selection techniques for pharmacokinetic analysis of dynamic contrast enhanced MRI in patients with glioblastoma tumors. *Medical Physics*, 43(6Part7):3392–3392.

Bagher-Ebadian, H., Jafari-Khouzani, K., Mitsias, P. D., et al. (2011). Predicting final extent of ischemic infarction using artificial neural network analysis of multi-parametric MRI in patients with stroke. *PloS one*, 6(8):e22626.

Brocks, D. R. and Hamdy, D. A. (2020). Bayesian estimation of pharmacokinetic parameters: an important component to include in the teaching of clinical pharmacokinetics and therapeutic drug monitoring. *Research in Pharmaceutical Sciences*, 15(6):503.

Choi, Y., Kim, D., Lee, S.-K., et al. (2015). The added prognostic value of preoperative dynamic contrast-enhanced MRI histogram analysis in patients with glioblastoma: analysis of overall and progression-free survival. *American Journal of Neuroradiology*, 36(12):2235–2241.

Clark, K., Vendt, B., Smith, K., et al. (2013). The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *Journal of Digital Imaging*, 26(6):1045–1057.

Dansirikul, C., Choi, M., and Duffull, S. B. (2005). Estimation of pharmacokinetic parameters from non-compartmental variables using Microsoft Excel®. *Computers in Biology and Medicine*, 35(5):389–403.

De Naeyer, D., De Deene, Y., Ceelen, W. P., et al. (2011). Precision analysis of kinetic modelling estimates in dynamic contrast enhanced MRI. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 24(2):51–66.

Dehkordi, A. N., Kamali-Asl, A., Ewing, J. R., et al. (2017). An adaptive model for rapid and direct estimation of extravascular extracellular space in dynamic contrast enhanced MRI studies. *NMR in Biomedicine*, 30(5):e3682.

Dehkordi, A. N., Sina, S., and Khodadadi, F. (2021). A comparison of deep learning and pharmacokinetic model selection methods in segmentation of high-grade glioma. *Frontiers in Biomedical Technologies*.

Donahue, J., Anne Hendricks, L., Guadarrama, S., et al. (2015). Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2625–2634.

Gill, A. B., Anandappa, G., Patterson, A. J., et al. (2015). The use of error-category mapping in pharmacokinetic model analysis of dynamic contrast-enhanced MRI data. *Magnetic Resonance Imaging*, 33(2):246–251.

Gravina, M., Marrone, S., Piantadosi, G., et al. (2019). 3TP-CNN: radiomics and deep learning for lesions classification in DCE-MRI. In *International Conference on Image Analysis and Processing*, pages 661–671. Springer.

Hsu, Y.-H. H., Ferl, G. Z., and Ng, C. M. (2013). GPU-accelerated nonparametric kinetic analysis of DCE-MRI data from glioblastoma patients treated with bevacizumab. *Magnetic Resonance Imaging*, 31(4):618–623.

Hỳvlová, D., Jiřík, R., and Vitouš, J. (2024). Deep-learning estimation of second-generation pharmacokinetic-model parameters in DCE-MRI. In *2024 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, pages 1–8. IEEE.

Jung, W., Bollmann, S., and Lee, J. (2022). Overview of quantitative susceptibility mapping using deep learning: current status, challenges and opportunities. *NMR in Biomedicine*, 35(4):e4292.

Kayastha, M. B., Liu, T., Titze, D., et al. (2023). Reconstructing 42 Years (1979–2020) of Great Lakes surface temperature through a deep learning approach. *Remote Sensing*, 15(17):4253.

Klepaczko, A., Strzelecki, M., Kociołek, M., et al. (2020). A multi-layer perceptron network for perfusion parameter estimation in DCE-MRI studies of the healthy kidney. *Applied Sciences*, 10(16):5525.

Li, X., Zhu, Y., Kang, H., et al. (2015). Glioma grading by microvascular permeability parameters derived from dynamic contrast-enhanced mri and intratumoral susceptibility signal on susceptibility weighted imaging. *Cancer Imaging*, 15(1):4.

Lonser, R. R., Sarntinoranont, M., and Bankiewicz, K. (2019). *Nervous system drug delivery: principles and practice*. Academic Press.

Metzler, C. M. (1986). Estimation of pharmacokinetic parameters: statistical considerations. *International Encyclopedia of Pharmacological Therapy*, pages 407–420.

NV, D. A. and Koohestani, S. (2019). The influence of signal to noise ratio on the pharmacokinetic analysis in dce-mri studies.

Oh, G., Moon, Y., Moon, W.-J., et al. (2024). Unpaired deep learning for pharmacokinetic parameter estimation from dynamic contrast-enhanced MRI without AIF measurements. *Neuroimage*, 291:120571.

Ottens, T., Barbieri, S., Orton, M. R., et al. (2022). Deep learning DCE-MRI parameter estimation: Application in pancreatic cancer. *Medical Image Analysis*, 80:102512.

Reavey-Cantwell, J. F., Haroun, R. I., Zahurak, M., et al. (2001). The prognostic value of tumor markers in patients with glioblastoma multiforme: analysis of 32 patients and review of the literature. *Journal of Neuro-Oncology*, 55(3):195–204.

Rygh, C. B., Wang, J., Thuen, M., et al. (2014). Dynamic contrast enhanced MRI detects early response to adoptive NK cellular immunotherapy targeting the NG2 proteoglycan in a rat model of glioblastoma. *PLoS One*, 9(9):e108414.

Schabel, M. C. and Parker, D. L. (2008). Uncertainty and bias in contrast concentration measurements using spoiled gradient echo pulse sequences. *Physics in Medicine & Biology*, 53(9):2345–2373.

Simeth, J. and Cao, Y. (2020). Gan and dual-input two-compartment model-based training of a neural network for robust quantification of contrast uptake rate in gadoxetic acid-enhanced MRI. *Medical Physics*, 47(4):1702–1712.

Tofts, P. S., Brix, G., Buckley, D. L., et al. (1999). Estimating kinetic parameters from dynamic contrast-enhanced T1-weighted MRI of a diffusable tracer: standardized quantities and symbols. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 10(3):223–232.

Tofts, P. S. and Kermode, A. G. (1991). Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts. *Magnetic Resonance in Medicine*, 17(2):357–367.

Ulas, C., Das, D., Thrippleton, M. J., et al. (2019). Convolutional neural networks for direct inference of pharmacokinetic parameters: application to stroke dynamic contrast-enhanced MRI. *Frontiers in Neurology*, 9:1147.

Ulas, C., Tetteh, G., Thrippleton, M. J., et al. (2018). 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*, pages 39–47. Springer.

Urien, S. and Lokiec, F. (2004). Population pharmacokinetics of total and unbound plasma cisplatin in adult patients. *British Journal of Clinical Pharmacology*, 57(6):756–763.

Wang, S., Liu, P., Turkbey, B., et al. (2012). Gaussian process inference for estimating pharmacokinetic parameters of dynamic contrast-enhanced MR images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 582–589. Springer.

Xue, X., Feng, J., Gao, Y., et al. (2019). Convolutional recurrent neural networks with a self-attention mechanism for personnel performance prediction. *Entropy*, 21(12):1227.

Yan, F.-R., Zhang, P., Liu, J.-L., et al. (2014). Parameter estimation of population pharmacokinetic models with stochastic differential equations: Implementation of an estimation algorithm. *Journal of Probability and Statistics*, 2014(1):836518.

Yu, R.-h. and Cao, Y.-x. (2017). A method to determine pharmacokinetic parameters based on andante constant-rate intravenous infusion. *Scientific Reports*, 7(1):13279.

Zhao, J., Huang, F., Lv, J., et al. (2020). Do RNN and LSTM have long memory? In *International Conference on Machine Learning*, pages 11365–11375. PMLR.

Ziayee, F., Müller-Lutz, A., Gross, J., et al. (2018). Influence of arterial input function (AIF) on quantitative prostate dynamic contrast-enhanced (DCE) MRI and zonal prostate anatomy. *Magnetic Resonance Imaging*, 53:28–33.

Zou, J., Balter, J. M., and Cao, Y. (2020). Estimation of pharmacokinetic parameters from DCE-MRI by extracting long and short time-dependent features using an LSTM network. *Medical Physics*, 47(8):3447–3457.

To cite this article:

A. NV Dehkordi, H. Malekmohammadi, S. Sina. PKest-LSnet: A robust LSTM-Based deep learning framework for accurate pharmacokinetic parameter estimation in DCE-MRI with clinical validation. *Radiation Physics and Engineering*, In Press.

DOI:

To link to this article: