

# Estimation of pharmacokinetic parameters from DCE-MRI by extracting long and short time-dependent features using an LSTM network

Jiaren Zou<sup>a)</sup>

*Department of Radiation Oncology, University of Michigan, Ann Arbor, MI 48109, USA*

James M. Balter

*Department of Radiation Oncology, University of Michigan, Ann Arbor, MI 48109, USA*

*Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109, USA*

Yue Cao

*Department of Radiation Oncology, University of Michigan, Ann Arbor, MI 48109, USA*

*Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI 48109, USA*

*Department of Radiology, University of Michigan, Ann Arbor, MI 48109, USA*

(Received 25 November 2019; revised 13 April 2020; accepted for publication 29 April 2020; published 3 June 2020)

**Purpose:**  $T_1$ -weighted dynamic contrast-enhanced Magnetic Resonance Imaging (DCE-MRI) is typically quantified by least squares (LS) fitting to a pharmacokinetic (PK) model to yield parameters of microvasculature and perfusion in normal and disease tissues. Such fitting is both time-consuming as well as subject to inaccuracy and instability in parameter estimates. Here, we propose a novel neural network approach to estimate the PK parameters by extracting long and short time-dependent features in DCE-MRI.

**Methods:** A Long Short-Term Memory (LSTM) network, widely used for processing sequence data, was employed to map DCE-MRI time-series accompanied with an arterial input function to parameters of the extended Tofts model. Head and neck DCE-MRI from 103 patients were used for training and testing the LSTM model. Arterial input functions (AIFs) from 78 patients were used to generate synthetic DCE-MRI time-series for training, during which data augmentation was used to overcome the limited size of *in vivo* data. The model was tested on independent synthesized DCE data using AIFs from 25 patients. The LSTM performance was optimized for the numbers of layers and hidden state features. The performance of the LSTM was tested for different temporal resolution, total acquisition time, and contrast-to-noise ratio (CNR), and compared to the conventional LS fitting and a CNN-based method.

**Results:** Compared to LS fitting, the LSTM model had comparable accuracy in PK parameter estimations from fully temporal-sampled DCE-MRI data (~3 s per frame), but much better accuracy for the data with temporally subsampling (4s or greater per frame), total acquisition time truncation by 48%–16%, or low CNR (5 and 10). The LSTM reduced normalized root mean squared error by 40.4%, 46.9%, and 53.0% for sampling intervals of 4s, 5s, and 6s, respectively, compared to LS fitting. Compared to the CNN model, the LSTM model reduced the error in the parameter estimates up to 55.2%. Also, the LSTM improved the inference time by ~ 14 times on CPU compared to LS fitting.

**Conclusion:** Our study suggests that the LSTM model could achieve improved robustness and computation speed for PK parameter estimation compared to LS fitting and the CNN based network, particularly for suboptimal data. © 2020 American Association of Physicists in Medicine [<https://doi.org/10.1002/mp.14222>]

Key words: contrast agent, DCE-MRI, long-short-term memory, machine learning, pharmacokinetic model, temporal correlation

## 1. INTRODUCTION

Dynamic contrast-enhanced Magnetic Resonance Imaging (DCE-MRI), an imaging technique where a time-series of  $T_1$  weighted images are obtained before, during, and after a bolus administration of a contrast agent (CA), has been widely explored in a wide range of clinical applications for noninvasive cancer detection, characterization, radiation target definition, and treatment response assessment.<sup>1–4</sup> The change in concentrations of CA in images over time can be derived from the MR signal intensity time-series. A

parametric pharmacokinetic (PK) model is then typically fit to the CA concentration-time curves to extract quantitative parameters related to vascular permeability, tissue perfusion, and volume of the extravascular extracellular space.<sup>5</sup>

Practical implementation of DCE-MRI involves a tradeoff of spatial resolution, volume of coverage, and temporal resolution. To estimate PK parameters accurately, an adequate temporal resolution of the DCE images is required, which often results in a tradeoff of spatial resolution and/or volume coverage. Several rapid imaging techniques have been introduced to improve both spatial resolution/volume coverage and

temporal resolution, e.g., k-space sparse sampling strategies combined with image reconstruction techniques of compressed sensing<sup>6</sup> and parallel imaging.<sup>7,8</sup> After reconstructing image time-series, the PK parameter maps are calculated as a second step, refereed as an indirect method. These methods include nonlinear least-squares (NLLSQ) fitting, which was reported to be sensitive to sampling interval, total acquisition time, and noise,<sup>9</sup> and requires intensive computation. A more efficient linear least-squares (LLSQ) method<sup>10</sup> was developed and showed improved accuracy for signals with signal-to-noise ratio less than 10. More recently, an efficient derivative based LLSQ method with a low-pass filter in time domain<sup>11</sup> was introduced. However, for the methods developed so far, obvious degradation of performance with increasing noise level and decreasing temporal resolution is still observed, and the computation time can be further improved. Recently, machine learning methods have been investigated to learn a mapping from fully sampled or subsampled image time-series to the parameter maps utilizing 2D or 3D convolutional neural networks (CNN).<sup>12,13</sup> Despite the short inference time, the CNN is not designed to learn long and short-term temporal relationships in the hemodynamics of the CA from the image time-series. Also, the CNN models may suffer from bias in practice since they were trained on similar subject arterial input functions (AIFs). In our experiment, where subject AIFs have diverse shapes, peaks, and time delays, obvious bias can be observed from the estimation from these models.<sup>13</sup>

Recurrent neural networks (RNN), especially long-short-term memory (LSTM),<sup>14</sup> have been successfully applied to learn temporal relationships in sequence data, such as video description and image captioning.<sup>15</sup> The application of LSTM was also explored in biomedical image analysis. A modified U-net was combined with an LSTM variant to do 3D biomedical volume segmentation,<sup>16</sup> where the LSTM explored the correlation of slices in the cranial-caudal direction. A pretrained fine-tuned Visual Geometry Group (VGG) network was used to extract feature maps from DCE-MRI slices, and then the sequence of feature maps was sent into an LSTM to determine whether a breast lesion within a given region of interest (ROI) was benign or malignant.<sup>17</sup>

Inspired by the recent successes of LSTM and the challenges faced by CNN-based approaches for PK parameter estimation, we propose a LSTM-based approach to learn the mapping of temporal dynamics in single-voxel signals accompanied, with their corresponding AIFs, to the PK parameters in the extended Tofts model.<sup>18</sup> Our approach is motivated by four factors. First, the signal intensity-time curves of DCE-MRI describe temporal hemodynamics of a CA passing through microvasculature in tissue. The LSTM architecture is able to learn long-term (temporal) dependence of signals<sup>15,19</sup> and thus could improve performance of PK parameter estimation compared to the CNN-based approaches using this data. Second, in practice, the AIF varies from patient to patient.<sup>20</sup> Inclusion of a subject-specific AIF in the estimation process could significantly improve estimation performance. AIFs can be readily incorporated into the input of LSTM as another input dimension, removing the bias observed in the reported

CNN-based approaches. Third, by capturing a low-dimensional manifold where the tissue concentration-time curve and AIF reside using LSTM, more robust parameter estimation can be achieved. Fourth, the inference time can be reduced because of the small computational burden of the LSTM as compared to DMF. We compare the results of LSTM with conventional direct model fitting<sup>21</sup> as well as a state-of-the-art CNN-based method,<sup>13</sup> including performances on the DCE-MRI at low contrast-to-noise ratio, low temporal sampling, and short total acquisition time.

## 2. MATERIALS AND METHODS

Our proposed LSTM-based method treats the PK parameter estimation problem as a mapping from a CA concentration-time curve accompanying with an AIF to the underlying parameters. Here, we investigated our method in the most commonly used extended Tofts model.

### 2.A. Extended Tofts model

The extended Tofts model<sup>18</sup> has been well-described in literature with a few minor variations in implementation. The version of the extended Tofts model and direct model fitting used in this study are described briefly. In our implementation, a contrast bolus arrival time (BAT),  $\tau_{BAT}(\mathbf{r})$  for each tissue voxel is considered for the accurate PK parameter estimation.<sup>22,23</sup> The implemented extended Tofts model is written as:

$$(1 - Hct)C_t(\mathbf{r}, t) = K^{trans}(\mathbf{r}) \int_0^t C_b(\tau - \tau_{BAT}(\mathbf{r}))e^{-K_{ep}(\mathbf{r})(t-\tau)}d\tau + v_p(\mathbf{r})C_b(t - \tau_{BAT}(\mathbf{r})) \quad (1)$$

where  $C_t(\mathbf{r}, t)$  is the CA concentration in the tissue voxel,  $C_b$  is the CA concentration in the arterial blood (or AIF),  $K^{trans}$  is the transfer rate constant of the CA that diffuses from the blood vessel to the interstitial space,  $K_{ep}$  is the rate constant of the CA efflux from the interstitial space to the blood plasma, and  $Hct$  is the small vessel hematocrit.  $K_{ep}$  equals to  $K^{trans}/v_e$ , where  $v_e$  is the fractional interstitial volume.

While [eq (1)] has been fit using NLLSQ methods<sup>10,11</sup> previously, in this investigation we selected a LLSQ method<sup>10</sup> that has a better tolerance to low SNR in the DCE data and a more efficient computation speed than NLLSQ fitting, as a benchmark to compare with our proposed LSTM method. In this LLSQ method, Eq [1] can be re-written as:

$$C_t(\mathbf{r}, t) = (K^{trans}(\mathbf{r}) + K_{ep}(\mathbf{r})v_p(\mathbf{r})) \int_0^t C_p(\tau - \tau_{BAT}(\mathbf{r}))d\tau - K_{ep}(\mathbf{r}) \int_0^t C_t(\mathbf{r}, \tau)d\tau + v_p(\mathbf{r})C_p(t - \tau_{BAT}(\mathbf{r})) \quad (2)$$

where  $K^{trans}$ ,  $K_{ep}$ , and  $v_p$  are linearly related to integrals of  $C_t(\mathbf{r}, t)$  and  $C_p(t)$ , and  $C_p(t)$ . For given  $\tau_{BAT}(\mathbf{r})$ ,  $K^{trans}$ ,  $K_{ep}$ , and  $v_p$  can be rapidly estimated by LLSQ fitting.  $\tau_{BAT}(\mathbf{r})$  can be estimated iteratively with  $K^{trans}$ ,  $K_{ep}$ , and  $v_p$ . The  $\tau_{BAT}(\mathbf{r})$  range can be determined using priori knowledge, e.g., 0-10 sec for the tissue in head and neck regions. In our implementation, we tested the  $\tau_{BAT}(\mathbf{r})$  between 0 and 10 sec with an incremental step 1s, and  $Hct$  as 0.45.<sup>9</sup> Hereafter, we refer this implementation of the LLSQ fitting as direct model fitting (DMF).

## 2.B. PK parameter inference via LSTM

### 2.B.1. Formulation

We estimate the PK parameters by mapping  $(C_t(\mathbf{r}, t), C_p(t))$  to the underlying physiological parameters  $\theta = (K^{trans}(\mathbf{r}), v_e(\mathbf{r}), v_p(\mathbf{r}))$  using LSTM, which we denote as  $\theta = f((C_t(\mathbf{r}, t), C_p(t)) | \Theta)$ , where  $f(\cdot | \Theta)$  represents the forward mapping of the LSTM network parameterized by  $\Theta$ .

### 2.B.2. Loss function

Our loss function seeks to reduce the mean squared error (MSE) between the estimated parameters  $\theta_e$  and the ground truth parameters  $\theta_g$  corresponding to the training signal series. Given a set of N training samples  $(c_{ti}(t), c_{pi}(t), \theta_{gi})$ ,  $i = 1, 2, \dots, N$ , we train the LSTM network to minimize the following loss function:

$$\mathcal{L}(\Theta) = \frac{1}{N} \sum_{i=1}^N \|\theta_{gi} - f((c_{ti}(t), c_{pi}(t)) | \Theta)\|_2^2 \quad (3)$$

### 2.B.3. LSTM network architecture

The proposed network (Fig. 1) consists of  $m$  LSTM layers with  $n$  sequentially connected cells in each layer. The network

takes an input sequence  $[C_t, C_p]^T$ , where the AIF is incorporated as another input dimension. The first LSTM layer extracts lower-level temporal relationships. The output feature sequence is then passed through the remaining  $m-1$  LSTM layers each to extract higher-order level temporal relationships from the signal and AIF.

Each LSTM layer captures the changes in the input sequence by maintaining a hidden state  $h^{(t)}$  and a memory cell  $c^{(t)}$  by updating them using gating mechanisms when stepping through the input sequence. Specifically, the  $l$  th LSTM layer takes a sequence of hidden states  $h_{l-1}^{(0)}, h_{l-1}^{(1)}, \dots, h_{l-1}^{(n-1)}$ ,  $l \in \{1, 2, \dots, m\}$ , where the superscript and subscript stand for timestep and layer, respectively, and  $h_0^{(0)}, h_0^{(1)}, \dots, h_0^{(n-1)}$  are defined as  $x^{(0)}, x^{(1)}, \dots, x^{(n-1)}$ . The new hidden states  $h_l^{(0)}, h_l^{(1)}, \dots, h_l^{(n-1)}$  are then defined by the equations shown in (4) below.

$$\begin{aligned} i_l^{(t)} &= \sigma(W_{in,i}h_{l-1}^{(t)} + W_{in,h}h_l^{(t-1)} + b_{i_l}) \\ f_l^{(t)} &= \sigma(W_{f,x}h_{l-1}^{(t)} + W_{f,h}h_l^{(t-1)} + b_{f_l}) \\ o_l^{(t)} &= \sigma(W_{o,x}h_{l-1}^{(t)} + W_{o,h}h_l^{(t-1)} + b_{o_l}) \\ \tilde{c}_l^{(t)} &= \tanh(W_{c,x}h_{l-1}^{(t)} + W_{c,h}h_l^{(t-1)} + b_{c_l}) \\ c_l^{(t)} &= \sigma(f_l^{(t)} \circ c_l^{(t-1)} + i_l^{(t)} \circ \tilde{c}_l^{(t)}) \\ h_l^{(t)} &= o_l^{(t)} \tanh(c_l^{(t)}) \end{aligned} \quad (4)$$

for  $t \in \{0, 1, \dots, n-1\}$  and  $l \in \{1, 2, \dots, m\}$ , where  $\sigma(\cdot)$  is the sigmoid function and “ $\circ$ ” denotes the Hadamard product.  $i_l^{(t)}$  and  $f_l^{(t)}$  control which information to “input” to and “forget” from the memory cell respectively. The new memory state  $c_l^{(t)}$  is then obtained based on the candidate values  $\tilde{c}_l^{(t)}$  and the gate values  $i_l^{(t)}$  and  $f_l^{(t)}$ . Finally, the new hidden state  $h_l^{(t)}$  is generated by the candidate values  $o_l^{(t)}$  and memory state  $c_l^{(t)}$ .

Batch normalization is applied after each LSTM layer except for the last one. A fully connected layer with three features is then applied to find the best combination of features to

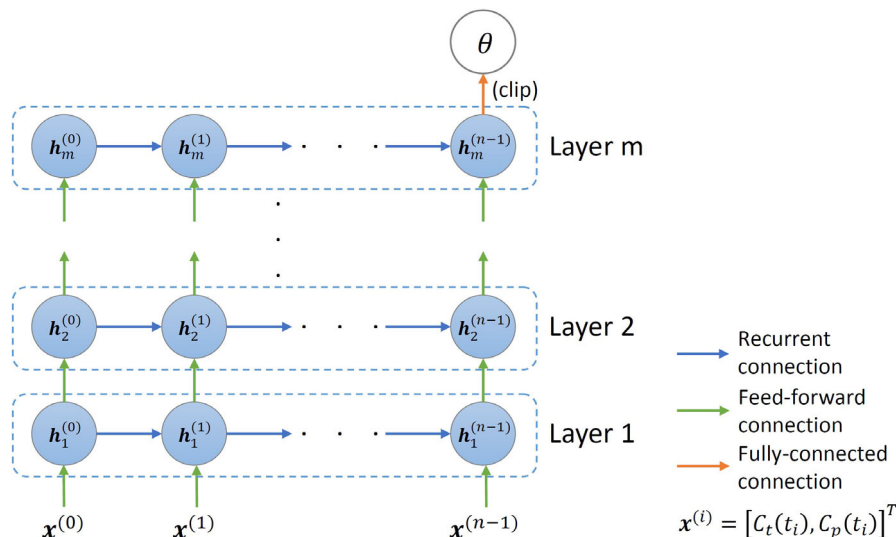


FIG. 1. Illustration of the network architecture used for PK parameter estimation from an input of a CA concentration time-series and an AIF as two separate channels. Each layer has  $n$  sequentially connected cells. [Color figure can be viewed at wileyonlinelibrary.com]

generate an estimation of the PK parameters. The parameters are clipped to our targeted range,  $K^{trans} \in [0, 3](min^{-1})$ ,  $v_e \in [0, 0.4]$ , and  $v_p \in [0, 0.55]$ , thus confining the parameters to fall within a physiologically realistic range,<sup>9</sup> to produce the final estimation of the PK parameters.

## 2.C. Data preparation

DCE MR time series for training and testing were synthesized using AIFs from 103 patients with head and neck cancers. As a brief, the DCE MRI images were acquired using a dynamic scanning sequence (TWIST) with an injection of 0.149 cc/kg of gadobenate dimeglumine on a 3 Tesla MRI scanner (Skyra, Siemens Healthineers, Erlangen Germany). The protocol was approved by the Institutional Review Board of the University of Michigan. The scanning parameters were: flip angle = 10°, echo time (TE) = 0.97 ms, repetition time (TR) = 2.73 ms, 60 time frames, voxel size =  $1.5625 \times 1.5625 \times 2.5 \text{ mm}^3$ , matrix =  $192 \times 192$ . There were small variations in time step of temporal sampling of the dynamic series (median of 3.34 sec) and in the numbers of the slices in z-direction between the patients (median of 72). For all cases, the subject-specific AIFs were extracted manually by averaging the signal intensity-time curves of 20 voxels from the carotid artery, which had maximum intensities at the time frame before the enhancement peak,<sup>24</sup> and then subtracting and dividing by the average precontrast signal intensities of the voxels. The targeted parameter maps were estimated using DMF.<sup>9</sup>

Of 103 patients, 78 cases were randomly selected for training, and 25 for testing. To overcome the limited size of the *in vivo* DCE-MRI dataset, synthetic data were created for network training and testing. The synthesized data allow us to obtain a reliable and accurate assessment of the performance of the proposed methods by comparing the estimates to the ground truth (the parameters that created the synthetic data). Data augmentation was also applied during training data synthesis. Using [eqs. (1)-(5)] of the extended Tofts model, the training signal intensity time-curves were created from different combinations of the AIF, time step ( $\cdot$  t), bolus arrival time ( $\tau_{BAT}$ ), and the parameters ( $\theta = (K^{trans}(\mathbf{r}), v_e(\mathbf{r}), v_p(\mathbf{r}))$ ). Particularly, AIFs and time steps of 78 training cases formed a set of 78 60-dimensional vectors and a set of 78 scalars, respectively, defined as  $A_{train}$  and  $T_{train}$ . This yielded a set of 75,678,643 3-dimensional parameter vectors ( $K^{trans}$ ,  $v_e$ , and  $v_p$ ), which was denoted as  $P_{train}$ . For testing, we used the DCE time-series synthesized using the AIFs from the 25 cases in the testing data pool as well as acquired empirical data. Testing with real data can assure that the model is 'realistic' enough. The synthesized testing data were generated in the same manner as the training data.

## 2.D. Experimental setup

Performances of the LSTM networks on DCE-MRI acquired with different total acquisition times and temporal sampling rates as well as different CNRs were compared with the DMF method and a CNN model.<sup>13</sup> The LSTM networks

were trained with input of fully temporal-sampled signal time-series as well as temporally subsampled time-series.

### 2.D.1. Training with synthetic data using acquired temporal sampling

The training data synthesis was executed on-the-fly during the network training and used the data from the training data generation pool consisting of the  $P_{train}$ ,  $A_{train}$ , and  $T_{train}$ . Figure 2 shows the data generating process during network training. Specifically, at each iteration, 1000 combinations of  $\{\theta_{g_i}\}_{i=1}^{1000}$ , AIFs, and time steps  $\{t\}_{i=1}^{1000}$ , were selected randomly from the aforementioned training data generation pool. For each combination, the AIF was first randomly scaled between 70% and 130% (AIF augmentation).<sup>25</sup> The concentration time-curve was then generated by randomly time-shifting the AIF (between 0 and 10 sec to simulate the delay of CA arrival) using the extended Tofts model, and random Gaussian noise was added to the signal time-series to have contrast-to-noise ratios between 20 and 30. The resultant signal time-series and the corresponding scaled AIF (without time shifting) were concatenated as an input. As a result, this batch data consisted of a vector of dimension  $1000 \times n \times 2$  (where n was the number of time points in the series), and was passed to the network. Each epoch consisted of 1000 batches of training data and 200 batches of validation data. The network was trained with Adam optimizer<sup>26</sup> with an initial learning rate of  $10^{-4}$ . The learning rate was reduced by a factor of 0.9 when the validation error was not improved in 30 consecutive epochs. The training was terminated when the validation error was not improved in 75 consecutive epochs.

To evaluate sufficiency of the size of the training data generation pool on performance of the LSTM model, two other models with the same architecture as that for the initial model (LSTM3) were trained with the same scheme using 60% (LSTM1) and 80% (LSTM2) of the training data generation pool.

### 2.D.2. Training with temporally subsampled synthetic data

The proposed LSTM network was also trained and tested on temporally subsampled synthetic signal time-curves, where signal time-curves were generated with sampling time steps of  $\Delta t$ : 3, 4, 5, and 6 sec. Lengths of the time-series were truncated at an integer number of time steps that was close to 168 sec. The AIF and the PK parameters were drawn from the same data generating pool. The LSTM input size was modified accordingly to match the number of time points of the subsampled signal time-series, while the other processes and parameters were kept the same, including the targeted PK parameters.

### 2.D.3. Training with truncated data and data with different CNR levels

To examine robustness of the proposed network to signal noise and total acquisition time, the proposed LSTM network



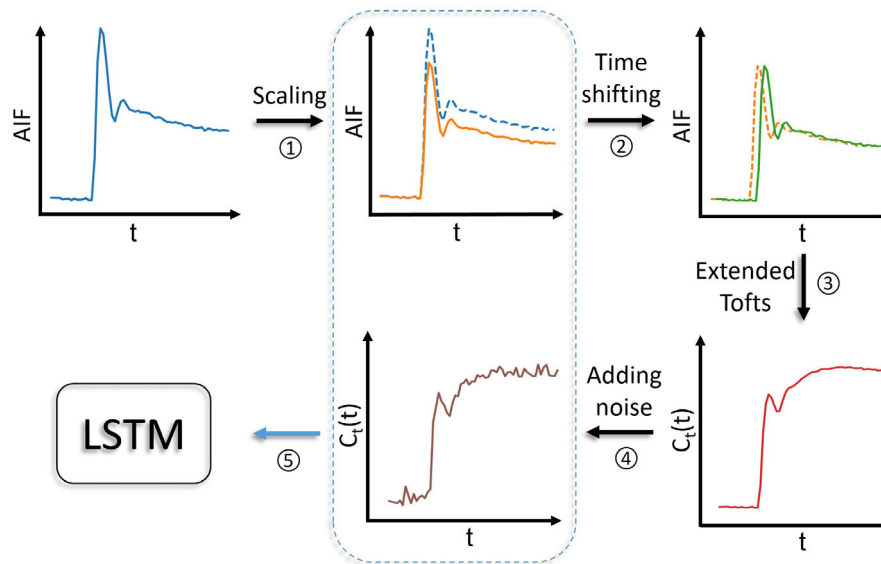


FIG. 2. Training data generating scheme with BAT simulation and AIF augmentation by random scaling. [Color figure can be viewed at wileyonlinelibrary.com]

was further trained and tested on: (a) truncated synthetic data with the total acquisition times 168s, 141s, 114s, and 87s when keeping the same sampling time step 3 sec, and (b) fully temporally sampled synthetic data but with CNR variations at 20–30, 10 and 5.

#### 2.D.4. Performance evaluation and comparison

First, our proposed LSTM network was optimized for the numbers of LSTM layers and hidden state features using fully temporal sampled synthetic data. Particularly, 2, 4, 6 and 8 layers, and 16, 32 and 64 features were trained and tested. Then, the optimal numbers of layers and features were used to train and test the LSTM model, whose performance was compared to both the conventional DMF approach and a CNN model. In testing, both synthetic and actual patient data were used. The synthetic testing data allows us to quantitatively assess the performance of PK parameters estimation by calculating the structural similarity (SSIM) index and normalized root mean squared error (NRMSE) of the estimated parameter maps with respect to the ground truth parameter maps. The NRMSE is defined by

$$NRMSE_i = \sqrt{\frac{\sum_{j=1}^N (\tilde{\theta}_{ij} - \theta_{ij})^2}{N-1}} / \left( \max_j \{\theta_{ij}\} - \min_j \{\theta_{ij}\} \right) \times 100(\%)$$

for a slice, where  $NRMSE_i$ ,  $\tilde{\theta}_{ij}$ ,  $\theta_{ij}$ , and  $N$  are the NRMSE of the  $i$ th parameter, the estimated and ground truth  $i$ th parameter for  $j$ th voxel, and the number of voxels in the slice. To synthesize the testing signal time-curves, the AIF and the PK parameter maps were from the same patients, for which no cross-combination of the AIF and the PK parameters nor AIF augmentation were used for the training data synthesis. The

SSIM and NRMSE were calculated from each of the 2D PK parameter slice maps first, and then averaged over multiple slices and across 25 cases. The proposed LSTM model was further evaluated using the empirical data from the same 25 cases.

The CNN model proposed by Ulas and colleagues<sup>13</sup> was implemented for comparison with the LSTM model. The CNN was trained on 1500 3D volumes with  $\ell_2$  loss and validated on 300 3D volumes generated using the same data generation pool as for LSTM training. The model loss term in the original model was dropped since BAT was not considered in the original paper, and enforcing model consistency without BAT correction produced worse results in our experiment. The size effect of the training data generation pool on the performance of the CNN-based method was also investigated by training with 60% (CNN1), 80% (CNN2), and 100% (CNN3) of the training data generation pool.

All codes were implemented using Keras library with Tensorflow backend, and the experiments were performed on an NVIDIA Tesla K40c GPU with 12 GB RAM.

## 3. RESULTS

### 3.A. Optimization of the LSTM networks

We investigated the distributions of the PK parameters  $\theta = (K^{trans}(\mathbf{r}), v_e(\mathbf{r}), v_p(\mathbf{r}))$  of the 78 patients, which were used as the training data generation pool. The parameter values in the pool covered the whole desired ranges of the parameters (Figure S1).

We trained the LSTM networks with 2, 4, 6, and 8 layers and 32 features, and with 16, 32, and 64 features and 6 layers. Figure 3 shows the impact of varying the layers and features on NRMSE% and SSIM. The LSTM network with 6 layers and 32 features had average maximal SSIM, presenting a balance of overfitting and underfitting of the training data and

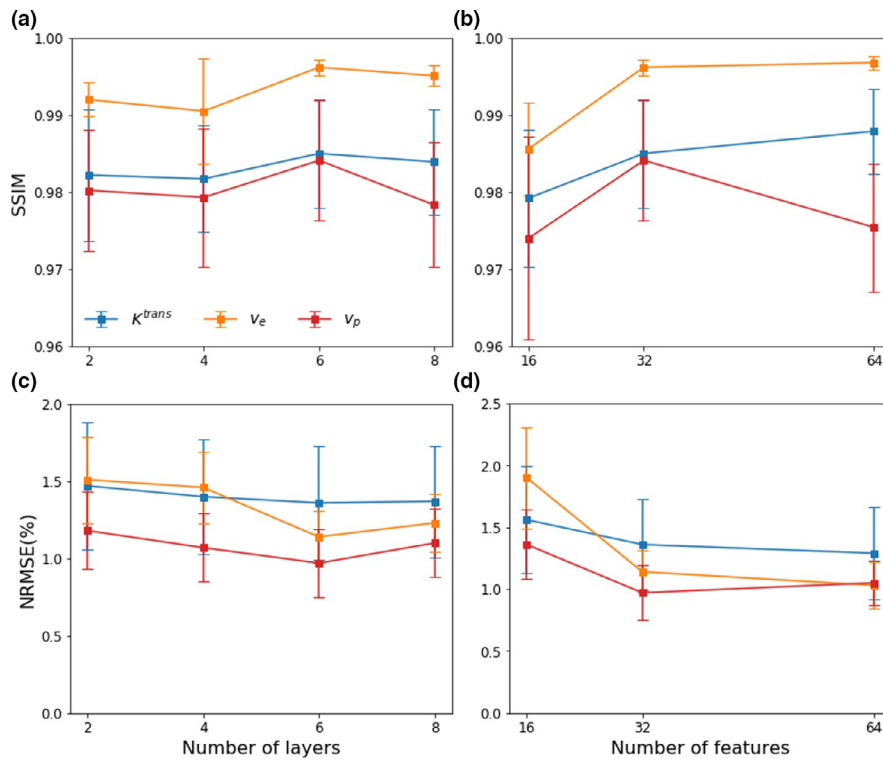


FIG. 3. SSIM and NRMSE of the parameter map estimation by LSTM networks with (a, c) 2, 4, 6, and 8 LSTM layers with 32 features, and (b, d) 16, 32, and 64 features with 6 layers. Error bar: standard deviation. [Color figure can be viewed at wileyonlinelibrary.com]

thus this network configuration was used for training under various test conditions.

### 3.B. Performance of the LSTM networks trained with original temporal-sampling data

The performance of the LSTM networks trained with synthetic data with original temporal-sampling, as well as results using the CNN and DMF models are shown in Table I. The proposed LSTM network achieved high SSIM and low NRMSE% in the whole field of view as well as in the gross tumor volume (Table I), comparable to the DMF approach. LSTM3 had < 1% lower SSIMs for  $K^{trans}$  and  $v_p$ , higher SSIM for  $v_e$ , and 13.4%, 19.1%, and 25.4% better NRMSEs for  $K^{trans}$ ,  $v_e$  and  $v_p$ , respectively, than the DMF approach. LSTM3 outperformed the CNN-based approach by reducing the NRMSE up to 55.2%. When evaluating the size effect of the data generation pool, the performance of LSTM1, LSTM2, and LSTM3 were similar, and the size effect was insignificant, indicating that the training data augmentation is effective. The CNN-based method shows inferior performance compared to the LSTM3 (Table I). A visual illustration of estimated PK parameters and residuals of the testing data by different methods is given in Figure S2.

Figure 4 shows the performance of the LSTM and CNN-based methods with different amounts of training data. The performance of CNN-based method changed little with an increase in the size of the training data generation pool, indicating the sufficiency of training data.

Figure 5 shows examples of the parameter maps generated by the DMF and LSTM methods on the acquired DCE-MRI data from the testing datasets, where the estimations of DMF and LSTM approaches were highly consistent. In most cases, DMF and LSTM approaches fit the signal intensity-time curves similarly well. Examples of fitted signal intensity-time curves are shown in Figure S3. To quantitatively compare the fitting results by LSTM and DMF, the voxel-wise MSE was calculated for both methods. The LSTM approach yielded a lower voxel-wise MSE ( $0.743 \pm 0.130$ ) than the DMF ( $0.808 \pm 0.127$ ). The parameter inference time of LSTM for fully temporal sampled DCE-MRI volumes was 250 s on CPU (40 s on GPU), while DMF approach required 3600 s to generate the PK parameter maps for the same data on the same CPU, representing approximately 14.4 times improvement in computation speed.

### 3.C. Performance of the LSTM networks with temporally subsampled data

Figure 6 and Table 2 show that the LSTM had significantly better performances for estimating parameters than the DMF method ( $P < 3.2 \times 10^{-6}$ ) when increasing temporal sampling intervals from 3s to 4s, 5s, and 6s (Figure S4).

### 3.D. Performance of the LSTM networks with low CNR and reduced total acquisition length

The performance of LSTM on DCE-MRI signals with lower CNRs and reduced total acquisition times is presented

TABLE I. Quantitative performance of different methods on test DCE-MRI volumes.

Method		SSIM			NRMSE (%)		
		$K^{trans}$	$v_e$	$v_p$	$K^{trans}$	$v_e$	$v_p$
Whole Range	DMF	0.9875 ± 0.0052	0.9960 ± 0.0010	0.9880 ± 0.0046	1.57 ± 0.43	1.41 ± 0.27	1.30 ± 0.38
	LSTM1	0.9853 ± 0.0068	0.9922 ± 0.0027	0.9806 ± 0.0159	1.47 ± 0.36	1.44 ± 0.25	1.15 ± 0.24
	LSTM2	0.9840 ± 0.0078	0.9931 ± 0.0025	0.9850 ± 0.0060	1.39 ± 0.41	1.19 ± 0.14	0.97 ± 0.18
	LSTM3	0.9850 ± 0.0070	0.9962 ± 0.0010	0.9841 ± 0.0078	<b>1.36 ± 0.37</b>	<b>1.14 ± 0.17</b>	<b>0.97 ± 0.22</b>
	CNN1	0.9547 ± 0.0202	0.9534 ± 0.0180	0.8946 ± 0.0267	3.07 ± 1.32	4.04 ± 1.57	3.54 ± 1.37
	CNN2	0.9629 ± 0.0159	0.9594 ± 0.0146	0.9323 ± 0.0237	2.78 ± 0.99	4.17 ± 1.32	3.16 ± 1.44
	CNN3	0.9627 ± 0.0189	0.9556 ± 0.0192	0.9302 ± 0.0352	2.76 ± 1.10	3.95 ± 1.30	2.85 ± 0.94
<i>P</i>	LSTM3 vs DMF	<0.05	0.54	<0.005	<0.05	<0.005	<0.005
<i>P</i>	LSTM3 vs CNN3	<0.005	<0.005	<0.005	<0.005	<0.005	<0.005
GTV	DMF	0.9994 ± 0.0006	0.9997 ± 0.0004	0.9994 ± 0.0008	3.85 ± 2.64	0.85 ± 0.35	0.74 ± 0.40
	LSTM3	0.9993 ± 0.0006	0.9998 ± 0.0003	0.9994 ± 0.0007	4.45 ± 2.64	0.85 ± 0.21	0.71 ± 0.24
<i>P</i>		0.38	<0.05	0.63	<0.05	1.00	0.45

The LSTM models were trained and tested using synthetic data with original temporal-sampling. The SSIM and NRMSE% (mean + std) with respect to the ground truth parameter maps were obtained in the whole field of view. The bold numbers indicate significant differences ( $P < 0.05$ ) between LSTM and DMF. GTV: gross tumor volume.

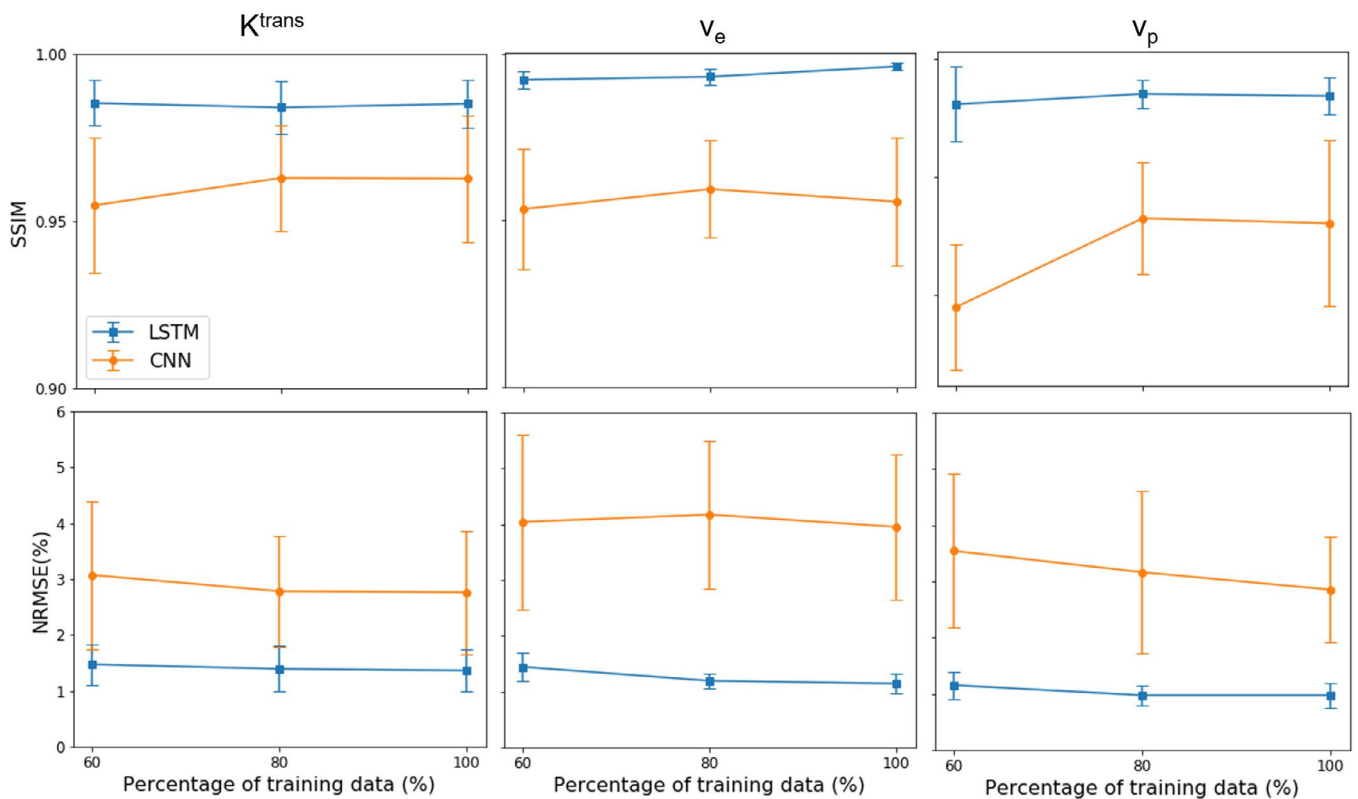


FIG. 4. The performance of the LSTM and CNN-based methods under different amounts of training data (60%, 80%, and 100% of the data generation pool). [Color figure can be viewed at wileyonlinelibrary.com]

in Figure 7 and Table S1. As can be seen, LSTM consistently improved the accuracy for lower CNR levels and reduced total acquisition times from 168s to 87s.  $v_e$  estimation had the largest improvement when total acquisition time was reduced.

### 4. DISCUSSION

We investigated a novel and potentially powerful LSTM-based network for learning a mapping from a CA

concentration-time curve including the corresponding AIF to the underlying PK parameters. The LSTM network is capable of learning long- and short-term dependency of sequence data such as DCE-MRI. We found that the performance of the LSTM on mapping DCE-MRI time-curves to their corresponding PK parameters was superior to a state-of-the-art CNN-based approach, and better than the direct model fitting method in terms of NRMSE. The LSTM was much more robust to temporally subsampled DCE data than the direct



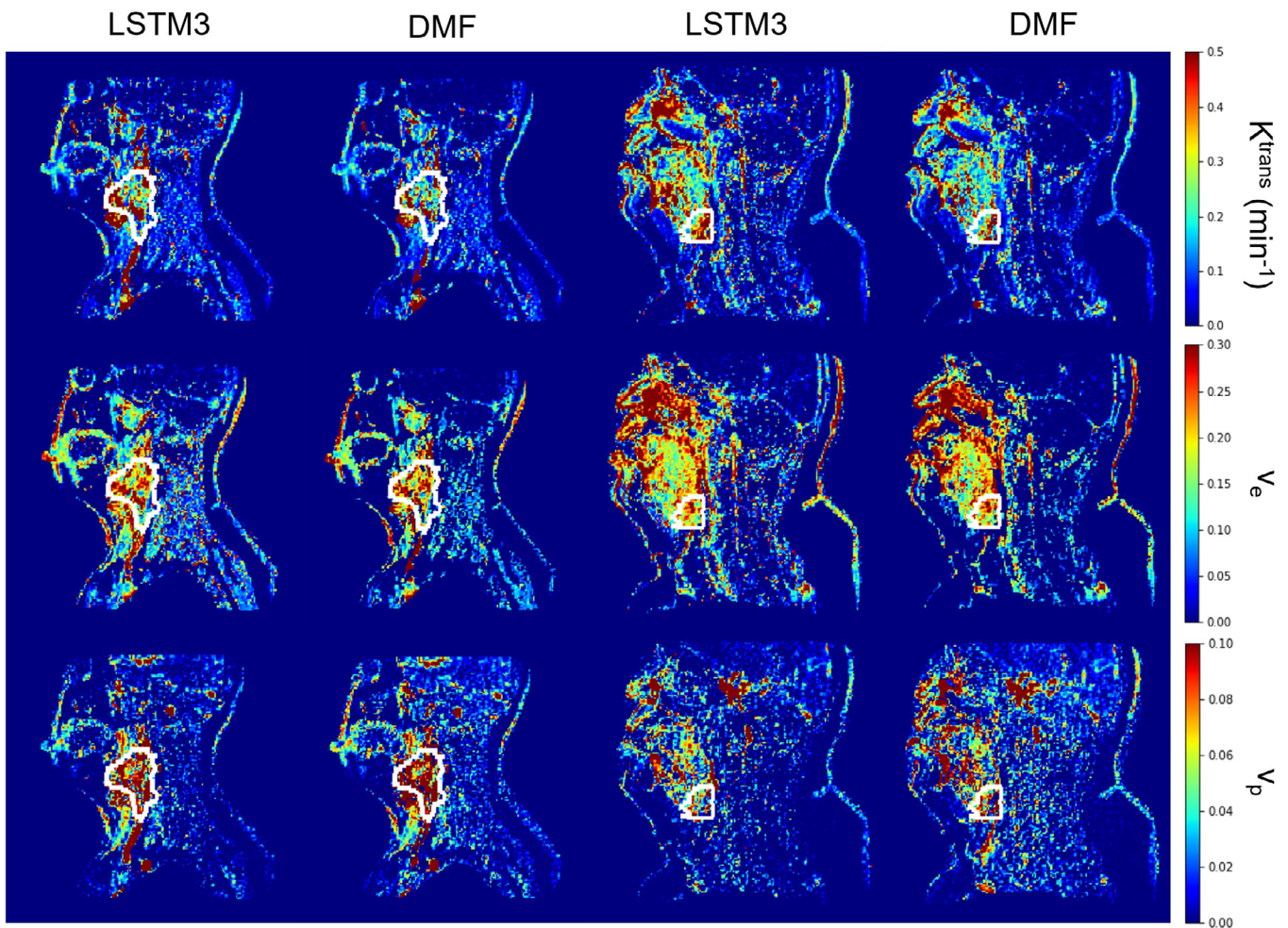


FIG. 5. Two exemplary slices (left 2 columns for first slice, right 2 columns for second slice) of PK parameter estimation by DMF and LSTM3 on an *in vivo* test dataset. The results obtained from DMF and LSTM3 show high similarity in both the tumor volume (depicted by the white contour) and the full FOV. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

PK model fitting, which can be utilized to increase spatial resolution of DCE images. Higher robustness of LSTM to noise and reduced acquisition time is also demonstrated compared with DMF. Our data augmentation strategies, including AIF augmentation and creation of synthetic signal time-curves from the data generation pool, overcame the limited size of the *in vivo* DCE-MRI training data pool. The LSTM network trained by the synthesized data was also able to perform well on empirical DCE data. This indicates that the synthetic data simulates real signal intensity-time series well and the LSTM network is effectively trained. In addition, our proposed network enables an approximately 90 times of computation time acceleration compared with the direct PK model fitting approach. The LSTM network has the potential to accelerate DCE-MRI acquisition and parameter estimation.

We attribute the superior performance of the LSTM to its capability to learn long- and short-term dependency in sequence data, and to extract dynamic features and temporal correlation in the signal intensity-time series of DCE-MRI. In contrast, the CNN extracts “spatial” features from the DCE-MRI volumes but has a limited capability to exploit the temporal relationships in the DCE data. For example, a 2D

CNN model treats the x-y-t data as a 3D volume, in which time-dependent features in the dynamic data could not be effectively extracted.<sup>27,28</sup> A 3D CNN model<sup>12</sup> attempts to address some of the issues of the 2D CNN by leveraging more temporal correlation. Without incorporating an AIF as input in the CNN, the PK parameters could have degeneracy and correspond to multiple signal time-curves, which can lead to a mis-mapping between the PK parameters and DCE curves. The small training and testing patch sizes limited by the high GPU memory demand of the 3D CNN training may further degrade its performance. Our proposed LSTM network structure is straightforward, and there is no PK model information required for LSTM training. Also, the AIF is incorporated into input as a second channel, which allows use of a patient-specific AIF when processing empirical patient DCE data. The proposed network can be easily trained and extended to other PK models, or even other sequence-related medical imaging data, e.g., a high-order diffusion model, with a minimum modification of the network architecture.

We used several strategies to overcome the limited size of the *in vivo* DCE-MRI data. First, we did not use a fixed triad



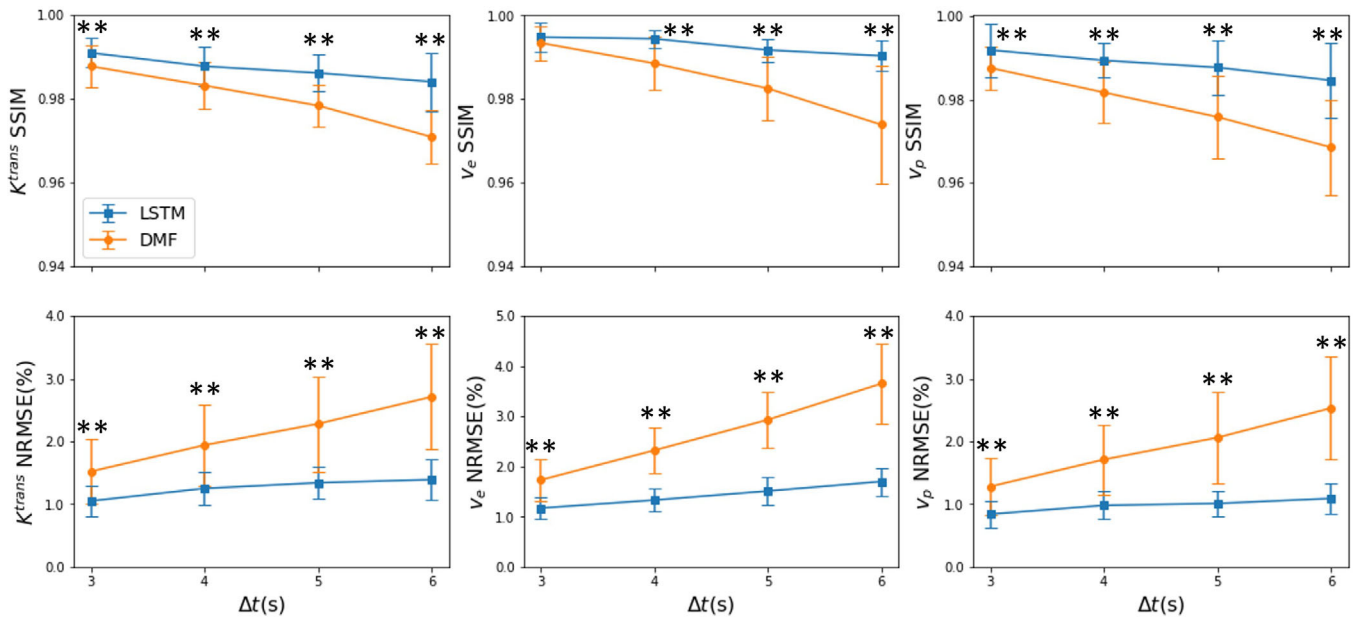


FIG. 6. Quantitative results of the estimated parameters from the 25 synthesized testing datasets with different temporal sampling time intervals (3, 4, 5, and 6s) by the LSTM and DMF approaches. The proposed LSTM shows a more stable performance than the DMF when increasing the sampling interval. Error bar: standard deviation.  $**P < 0.005$ . [Color figure can be viewed at wileyonlinelibrary.com]

TABLE II. Quantitative performance of different methods on temporally subsampled DCE data.

Method	SSIM			NRMSE (%)		
	$K^{trans}$	$V_e$	$V_p$	$K^{trans}$	$V_e$	$V_p$
$\Delta t = 3$						
DMF	0.9877 ± 0.0050	0.9934 ± 0.0041	0.9876 ± 0.0052	1.52 ± 0.52	1.73 ± 0.41	1.28 ± 0.46
LSTM3	<b>0.9909 ± 0.0035</b>	0.9948 ± 0.0034	<b>0.9919 ± 0.0064</b>	<b>1.05 ± 0.24</b>	<b>1.17 ± 0.21</b>	<b>0.84 ± 0.21</b>
$P$	<0.005	0.12	<0.005	<0.005	<0.005	<0.005
$\Delta t = 4$						
DMF	0.9831 ± 0.0055	0.9885 ± 0.0064	0.9817 ± 0.0073	1.94 ± 0.64	2.32 ± 0.46	1.71 ± 0.56
LSTM3	<b>0.9877 ± 0.0048</b>	<b>0.9944 ± 0.0022</b>	<b>0.9894 ± 0.0041</b>	<b>1.25 ± 0.27</b>	<b>1.33 ± 0.23</b>	<b>0.98 ± 0.22</b>
$p$	<0.005	<0.005	<0.005	<0.005	<0.005	<0.005
$\Delta t = 5$						
DMF	0.9783 ± 0.0050	0.9825 ± 0.0075	0.9758 ± 0.0099	2.28 ± 0.76	2.93 ± 0.56	2.06 ± 0.73
LSTM3	<b>0.9861 ± 0.0044</b>	<b>0.9917 ± 0.0027</b>	<b>0.9877 ± 0.0066</b>	<b>1.34 ± 0.26</b>	<b>1.51 ± 0.27</b>	<b>1.01 ± 0.21</b>
$P$	<0.005	<0.005	<0.005	<0.005	<0.005	<0.005
$\Delta t = 6$						
DMF	0.9708 ± 0.0063	0.9738 ± 0.0142	0.9685 ± 0.0115	2.71 ± 0.84	3.65 ± 0.80	2.53 ± 0.82
LSTM3	<b>0.9840 ± 0.0070</b>	<b>0.9903 ± 0.0037</b>	<b>0.9846 ± 0.0089</b>	<b>1.39 ± 0.33</b>	<b>1.70 ± 0.28</b>	<b>1.09 ± 0.25</b>
$P$	<0.005	<0.005	<0.005	<0.005	<0.005	<0.005

The SSIM and NRMSE% (mean + std) with respect to the ground truth parameter maps were obtained in the whole field of view. The bold numbers indicate significant differences ( $P < 0.005$ ) between the two methods.

of the AIF, the PK parameters, and the signal time-curves from the *in vivo* DCE-MRI dataset, which limits us to a total of 103 patient datasets. We used synthesized signal time-curves that were created by randomly selected and combined PK parameters, AIF (and augmented AIF), and time step from a data pool. In addition, we added random variations of the delay of CA bolus arrival into the signal time-curves. Our evaluation shows that the PK parameters from 78 patients sufficiently cover the parameter ranges of interest. The performance of LSTM2, and LSTM3 indicates that the PK parameters from 60% of the training data could sufficiently cover the ranges of parameter of interest. Our overall strategy of the training data synthesis seems to yield effective training

of the LSTM networks, and overcomes the limited size of *in vivo* datasets. These strategies seem to reasonably mitigate small amount of variation in sampling interval in the *in vivo* data.

For extraction of PK parameters from signal time curves with different time steps (3–6 sec), the LSTM models show robust performance compared to the direct PK model fitting, with the latter showing performance degradation with an increase in the time step size. This advantage of the LSTM approach can be utilized to improve the spatial resolution of DCE-MRI when decreasing the temporal resolution of DCE-MRI. For example, a modest increase in the spatial resolution from the currently used  $1.6 \times 1.6 \times 2.5 \text{ mm}^3$  to

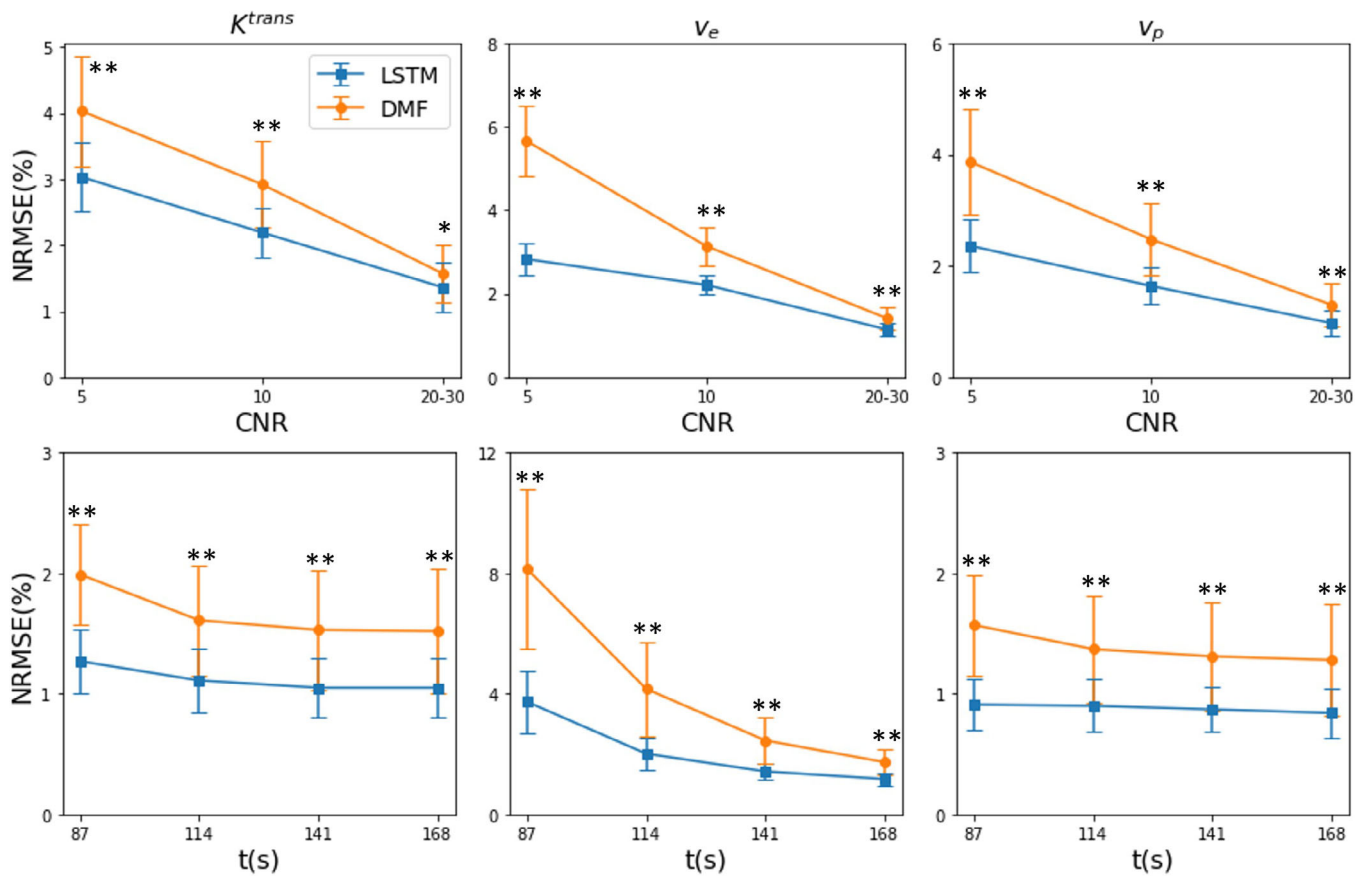


FIG. 7. The performance (NRMSE) of the LSTM and DMF estimations under lower CNRs (first row) with full temporal sampling and reduced total acquisition times  $t$  (second row) with  $\Delta t = 3$  s and  $CNR = 20-30$ . Error bar: standard deviation.  $*P < 0.05$ ;  $**P < 0.005$ . [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

$1.4 \times 1.4 \times 1.4 \text{ mm}^3$  would prolong the acquisition time of an image volume by a factor of 2. This increase in the temporal resolution from 3 sec to 6 sec would result in an increase in NRMSE% of the estimates by 1.7–2.1 times by the DMF but a very small increase by the LSTM model (Fig. 6).

This work has several limitations. As we can see from Fig. 4 and Table I, the LSTM performs better than DMF in terms of NRMSE but not as good in terms of SSIM across the whole parameter range. Further analysis of the error distribution reveals that this is mainly due to the minor estimation errors in parameter combinations of zero  $K^{trans}$  and  $v_e$  but nonzero  $v_p$  (mainly in the brain region), which is caused by the small portion of these parameter value combinations in the training datasets. Another concern is the accuracy of parameter estimation in the gross tumor volume, which has a different range of the parameters from normal tissue and has a small amount of the data weighting in the training dataset. We note similar SSIM values of DMF and LSTM in the gross tumor volume (Table I). Further manipulation of the training dataset distribution and/or a modification of the loss function or a weighting training data in different parameter ranges could improve performance of the algorithm. Another path to improvement generalizability is to incorporate temporal sampling intervals into the network input as another channel.<sup>29</sup> The performance of the current vanilla LSTM architecture

could be further improved by using a bidirectional LSTM with attention at the expense of longer training and inference time. A further improved model could use more realistic synthetic data that takes motion artifacts and other factors into account to improve the robustness of performance of the LSTM on *in vivo* DCE-MRI datasets.

## 5. CONCLUSIONS

In conclusion, our proposed LSTM is a promising approach to estimate PK parameters from DCE-MRI time-series. We demonstrate that the proposed approach provides more accurate PK parameter maps compared to the CNN-based approach, and is comparable to the DMF method with approximately 90 times of computation time reduction. The LSTM networks are more robust to low temporal resolution, lower CNR levels, and reduced total acquisition time than direct PK model fitting. In the future, a similar LSTM network can be applied directly to  $(k, t)$  space data to leverage the temporal and spatial sparsity, and thus accelerate both the data acquisition and image reconstruction of DCE-MRI.

## ACKNOWLEDGMENTS

This work was supported by NIH R01 EB016079.

## CONFLICT OF INTEREST

The authors have no conflicts to disclose.

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: jiarenz@med.umich.edu; Telephone: +1-734-800-6786

## REFERENCES

- Cao Y, Popovtzer A, Li D, et al. Early prediction of outcome in advanced head-and-neck cancer based on Tumor blood volume alterations during therapy: a prospective study. *Int J Radiat Oncol Biol Phys.* 2008;72:1287–1290.
- Gaddikeri S, Gaddikeri RS, Taylor T, Anzai Y. Dynamic contrast-enhanced MR imaging in head and neck cancer: techniques and clinical applications. *Am J Neuroradiol.* 2016;37:588–595.
- Cao Y, Aryal M, Li P, et al. Predictive values of MRI and PET derived quantitative parameters for patterns of failure in both p16+ and p16–High risk head and neck cancer. *Front Oncol.* 2019;9:1–11.
- Kim MM, Parmar HA, Aryal MP, et al. Developing a pipeline for Multi-parametric MRI-guided radiation therapy: initial results from a phase II clinical trial in newly diagnosed Glioblastoma. *Tomogr (Ann Arbor, Mich).* 2019;5:118–126.
- Khalifa F, Soliman A, El-Baz A, et al. Models and methods for analyzing DCE-MRI: a review. *Med Phys.* 2014;41:1–32.
- Lustig M, Donoho D, Pauly JM. Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn Reson Med.* 2007;58:1182–1195.
- Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Magn Reson Med.* 1999;42:952–962.
- Uecker M, Lai P, Murphy MJ, et al. ESPIRiT - An eigenvalue approach to autocalibrating parallel MRI: where SENSE meets GRAPPA. *Magn Reson Med.* 2014;71:990–1001.
- Cao Y, Li D, Shen Z, Normolle D. Sensitivity of quantitative metrics derived from DCE MRI and a pharmacokinetic model to image quality and acquisition parameters. *Acad Radiol.* 2010;17:468–478.
- Murase K. Efficient method for calculating kinetic parameters using T1-weighted dynamic contrast-enhanced magnetic resonance imaging. *Magn Reson Med.* 2004;51:858–862.
- Wang C, Yin FF, Chang Z. An efficient calculation method for pharmacokinetic parameters in brain permeability study using dynamic contrast-enhanced MRI. *Magn Reson Med.* 2016;75:739–749.
- Ulas C, Tetteh G, Thrippleton MJ, et al. Direct Estimation of Pharmacokinetic Parameters from DCE-MRI using Deep CNN with Forward Physical Model Loss. 2018;11070 LNCS:39-47.
- 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.* 2019;9.
- Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* 1997;9:1735–1780.
- Donahue J, Hendricks LA, Rohrbach M, et al. Long-term recurrent convolutional networks for visual recognition and description. *IEEE Trans Pattern Anal Mach Intell.* 2017;39:677–691.
- Chen J, Yang L, Zhang Y, Alber M, Chen DZ. Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation. In: *Advances in Neural Information Processing Systems.*; 2016:3044–3052. <http://arxiv.org/abs/1609.01006>.
- Antropova N, Huynh B, Li H, Giger ML. Breast lesion classification based on dynamic contrast-enhanced magnetic resonance images sequences with long short-term memory networks. *J Med Imaging.* 2018;6:1.
- Tofts PS. Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. *J Magn Reson Imaging.* 1997;7:91–101.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* 2015;521:436–444.
- Paldino MJ, Barboriak DP. Fundamentals of quantitative dynamic contrast-enhanced MR imaging. *Magn Reson Imaging Clin N Am.* 2009;17:277–289.
- Cao Y, WE-D-T- 6C-03: development of image software tools for radiation therapy assessment. *Med Phys.* 2005;32:2136–2136.
- Singh A, Rathore RKS, Haris M, Verma SK, Husain N, Gupta RK. Improved bolus arrival time and arterial input function estimation for tracer kinetic analysis in DCE-MRI. *J Magn Reson Imaging.* 2009;29:166–176.
- Bendinger AL, Debus C, Glowa C, Karger CP, Peter J, Storath M. Bolus arrival time estimation in dynamic contrast-enhanced magnetic resonance imaging of small animals based on spline models. *Phys Med Biol.* 2019;64:045003.
- Fritz-Hansen T, Rostrup E, Larsson HBW, Søndergaard L, Ring P, Henriksen O. Measurement of the arterial concentration of Gd-DTPA using MRI: a step toward quantitative perfusion imaging. *Magn Reson Med.* 1996;36:225–231.
- Robben D, Suetens P. Perfusion parameter estimation using neural networks and data augmentation. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* Vol 11383 LNCS. Springer Verlag; 2019:439-446.
- Kingma DP, Adam B. Adam: A method for stochastic optimization. In: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings.* 2015. <http://arxiv.org/abs/1412.6980>
- Arbelle A, Raviv TR. Microscopy cell segmentation via convolutional LSTM networks. In: *Proceedings - International Symposium on Biomedical Imaging.* Vol 2019-April. IEEE; 2019:1008-1012
- Zhang D, Icke I, Dogdas B, et al. A multi-level convolutional LSTM model for the segmentation of left ventricle myocardium in infarcted porcine cine MR images. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018).* Vol 2018-April. IEEE; 2018:470-473
- Simeth J, Cao Y. GAN and dual-input two-compartment model-based training of a neural network for robust quantification of contrast uptake rate in gadoteric acid-enhanced MRI. *Med Phys.* 2020;1–11.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Fig S1.** The distribution of (a) Ktrans, (b) ve, and (c) vp in the training data generation pool.

**Fig S2.** An exemplary slice of the ground truth parameter maps (column 1), the estimated maps (column 2-4), and the residual maps (last 3 columns) of estimated Ktrans (top row) ve (middle row) and vp (bottom row) by the LSTM3, CNN, and DMF models from a testing case. The white contour depicts the gross tumor volume. GT: ground truth.

**Fig S3.** Three exemplary in vivo CA concentration-time curve fitting results by DMF and LSTM3 in the tumor region indicated by the white contour shown in the left Ktrans parameter map estimation by DMF. Results from both methods are in reasonable alignment with observed data. (SE = squared error).

**Fig S4.** An exemplary slice of ground truth of parameter maps (column 1), estimated maps (column 2 and 3), and the residual maps (last 2 columns) by LSTM and DMF using temporal sampling interval  $\Delta t = 6s$ . The tumor volume is depicted by a white contour. GT: ground truth.

**Table S1.** NRMSE (%) of testing signals with lower CNR levels and reduced total acquisition time (t). The bold numbers are the better results of the proposed method than those of DMF.