

Estimation of the capillary level input function for dynamic contrast-enhanced MRI of the brain tumor using Vision Transformer approach

Author: Ximeng Zhang

Supervisor: Zhengguo Tan

1 Introduction

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) has emerged as a powerful tool for assessing tissue perfusion and vascular permeability[1], thereby enabling the quantitative analysis of tumor microenvironment and response to therapy. The pharmacokinetic modeling of DCE-MRI data is a critical component of this analysis, providing insights into physiological and pathological processes. The purpose of this article is to conduct an in-depth study of the pharmacokinetic model of DCE-MRI and use the Vision Transformer method to estimate the Estimation of the capillary level input function.

The pharmacokinetic model of DCE-MRI relies on precise measurement of AIF[2], which represents Adequate arterial input function as the input to the model. The AIF serves as the key input to the pharmacokinetic model, enabling the quantification of parameters such as perfusion, permeability, and volume of distribution. However, there are significant challenges associated with obtaining a reliable AIF, which have hindered the precision and reproducibility of DCE-MRI-based measurements[3].

One of the primary problems encountered in AIF measurement is the variability arising from the choice of the location where it is obtained. Measurement of the AIF from different anatomical sites, even within the same subject, often results in significant differences due to variations in local blood flow and vessel geometry. This can lead to inaccuracies in the pharmacokinetic model and the derived parameters[4].

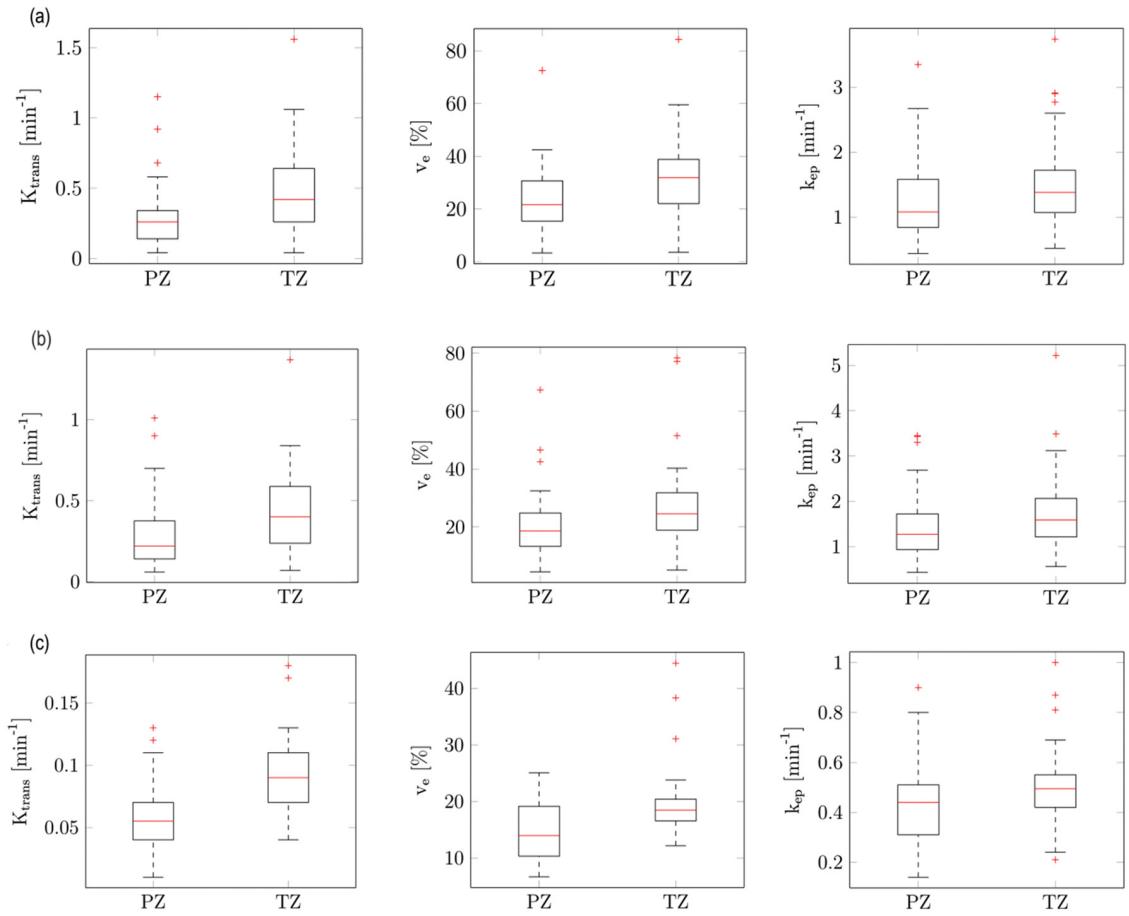


Figure 1: (a) Manually selected case-specific AIF. (b) Auto-selected case-specific AIF. (c) Population based AIF.[4]

Another issue that has been widely recognized is the operator dependence in AIF measurement. AIF obtained at the same site can yield different results when measured by different operators due to variations in technique and expertise. This operator-induced variability further compromises the reliability and reproducibility of DCE-MRI data.

To address these challenges, this article introduces a method for estimating pharmacokinetic models utilizing capillary-level input functions (CIFs) as an alternative to traditional AIFs. This method aims to improve the accuracy and reliability of DCE-MRI pharmacokinetic models by capturing the capillary-level input function (CIF) through a vision transformer neural network, which is expected to alleviate issues related to AIF variability and operator dependence. This approach promises insights into tissue heterogeneity, vascular patterns, and perfusion characteristics at a finer scale than previously possible.[5]

2 Methods

2.1 Dynamic contrast-enhanced MRI

Conventional Contrast-enhanced MRI has long been a valuable tool in medical imaging, offering critical structural insights into various tissues. However, it has inherent limitations as it primarily provides anatomical details and lacks functional information. This conventional approach typically offers a single static image captured shortly after the administration of a contrast agent [3]. In contrast, Dynamic Contrast-enhanced MRI (DCE-MRI) emerges as a unique and indispensable technique in the realm of medical imaging. DCE-MRI stands alone as the sole MRI method capable of evaluating tumors in terms of their state of functional microcirculation. It has become widely adopted for the diagnosis and staging of cancer, enabling the simultaneous depiction of both physiologic alterations and morphologic changes [2][3].

Dynamic contrast-enhanced MRI (DCE-MRI) is a powerful imaging modality that excels in capturing the perfusion characteristics of targeted tissues. By acquiring a series of dynamically acquired T1-weighted images with the introduction of a contrast agent, DCE-MRI enables the visualization of real-time changes in tissue perfusion. This dynamic imaging data is subsequently subjected to pharmacokinetic analysis, which yields valuable kinetic parameters providing insights into blood flow, vessel permeability, and tissue vascularization. Therefore, there is a growing need for a sensitive DCE-MRI technique capable of detecting even the most subtle changes in BBB permeability.

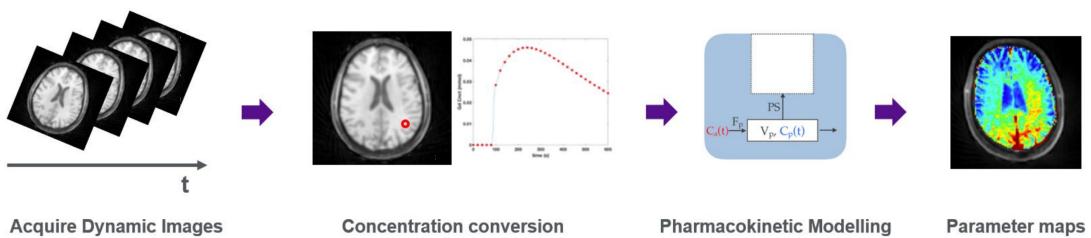


Figure 2: Kinetic parameter estimation process

2.2 Pharmacokinetic Model

Conventional pharmacokinetic model analysis of breast dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) data plays a pivotal role in the assessment of breast lesions and the characterization of their vascular properties. Several established models are routinely employed for this purpose, including the extended Tofts model

(eTofts), the two-compartment exchange model (TCM), and the extended vascular tree model (EVM). These models are instrumental in quantifying parameters related to tissue perfusion, vascular permeability, and blood volume, all of which are critical for distinguishing between benign and malignant breast lesions, as well as for monitoring treatment responses. The eTofts model focuses on characterizing vascular properties through the analysis of contrast agent kinetics in tissue. In contrast, the TCM delves deeper into tissue microstructure, accounting for the exchange of contrast agents between blood vessels and interstitial spaces. The EVM, on the other hand, provides an even more comprehensive depiction of vascular behavior by considering a broader range of vessel structures. The integration of these models in breast DCE-MRI analysis equips clinicians and researchers with a valuable toolkit for the precise and detailed evaluation of breast lesions and their vascular characteristics[3].

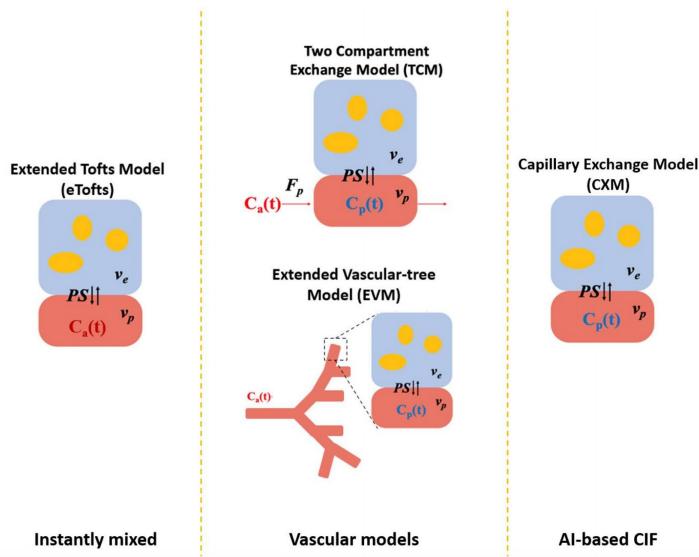


Figure 3: Illustrates the pharmacokinetic models employed in this study. On the left, the eTofts model depicts the rapid mixing of the arterial input function (AIF) within the capillary bed. In the middle, the Two-Compartment Exchange Model (TCM) and the Extended Vascular Tree Model (EVM) account for the intricate transport and dispersion of the AIF as it traverses to the capillary bed. On the right, the Contrast Transfer Model (CXM) integrates a capillary input function (CIF) predicted by a trained neural network as its input. Key estimated kinetic parameters across these models include v_e (volume fraction of extracellular-extravascular space), v_p (volume fraction of the blood plasma compartment), F_p (blood flow from the artery to the capillary bed), PS (bidirectional endothelial permeability-surface product), and t_0 (the time it takes for a contrast agent to traverse each branch of the arterial tree). These models provide a comprehensive framework for analyzing dynamic contrast-enhanced MRI data and extracting essential information related to tissue perfusion and vascular characteristics[3].

2.3 Vision Transformer neural network for estimating CIF

The prediction of Capillary Input Functions (CIF) using deep learning is an innovative approach that has the potential to revolutionize the field of dynamic contrast-enhanced imaging. In this method, a well-defined patch design is employed, ensuring that the local input function remains consistent throughout the dataset. Tissue concentration-time curves (Cts) are simulated based on kinetic parameters using the arterial input function (AIF) as a foundation. Deep learning networks are then employed to predict CIFs. The training process involves exposing the network to a diverse set of AIFs, allowing it to learn and adapt to the various shapes and patterns of tissue concentration-time curves.

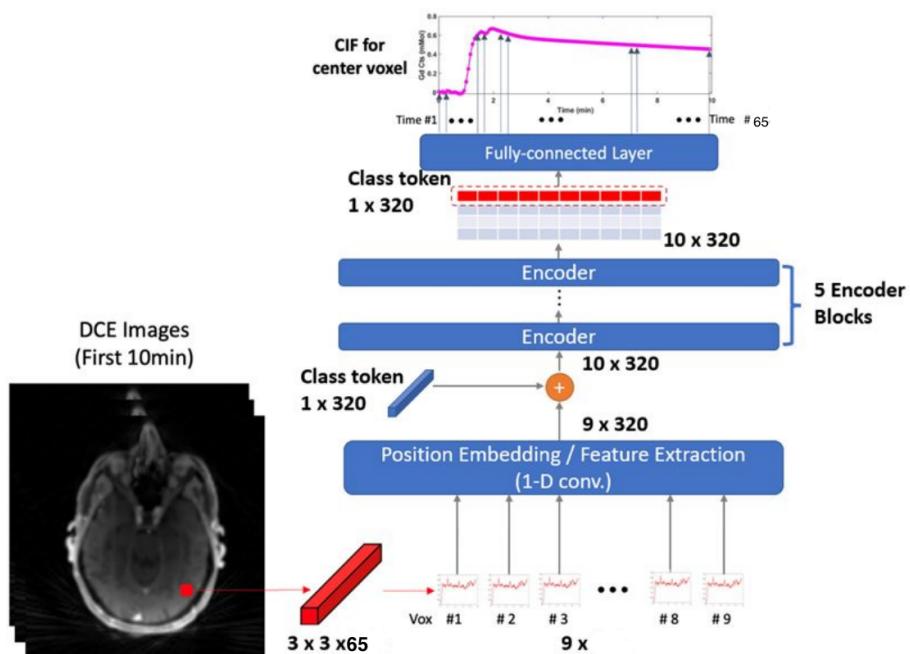


Figure 4: Main name 2

The Vision Transformer neural network (ViT) is a cutting-edge architecture that is increasingly being applied to diverse image analysis tasks. When applied to Dynamic Contrast-Enhanced (DCE) images in a $3 \times 3 \times 65$ format, the ViT first divides the image into nine distinct patches, effectively breaking down the input data into smaller, manageable components. This division is followed by the introduction of position embedding and feature extraction to help the network understand the spatial relationship of these patches. The result is a set of class tokens in a 1×320 format.

these class tokens through five layers of transformer encoders. These encoder layers facilitate contextual learning, allowing the network to discern intricate patterns and

relationships within the DCE data. The ultimate goal of this process is to estimate the Capillary Input Function (CIF) for the center voxel within the 65 records of the DCE image sequence. As each class token moves through a fully-connected layer, it undergoes transformations to produce an accurate estimation of the CIF, providing valuable insights into the microvascular dynamics within the targeted tissue region. This approach exemplifies the ViT's capability to effectively analyze complex medical imaging data and extract functional information.

2.4 Application to the RidersNeuro dataset

In our research endeavor, we undertook the task of developing and training a sophisticated Capillary Input Function (CIF) network with the capacity to adapt and accommodate the intricate and diverse concentration-time curves (Cts) observed within the complex landscape of brain tissues. These Cts vary considerably, particularly in cases involving brain tumors characterized by varying perfusion levels, as well as in the healthy normal brain tissue.

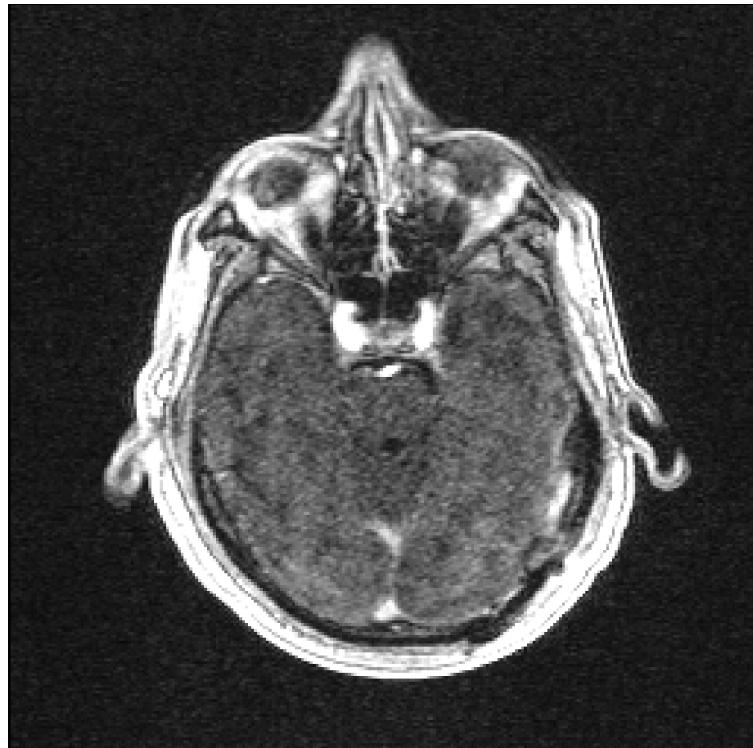


Figure 5: The structure of Vision Transformer

To comprehensively assess the performance of our CIF network, we designed a rigorous testing procedure aimed at determining the repeatability and reliability of CIF measurements. This evaluation was carried out using the rich and informative RidersNeuro

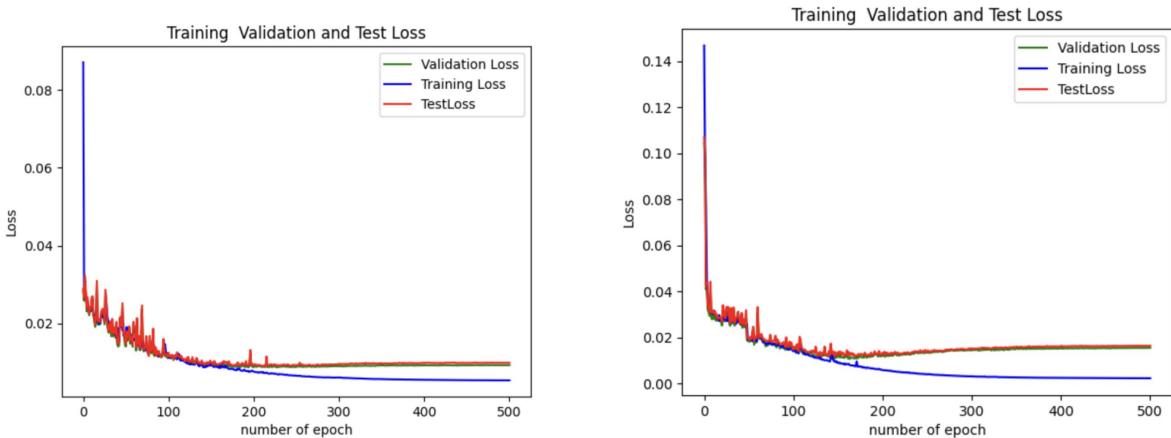
dataset, which included data from 19 patients diagnosed with brain tumors, with each patient undergoing two separate visits for an in-depth examination. The dataset was characterized by high-resolution Dynamic Contrast-Enhanced Magnetic Resonance Imaging (DCE-MRI) images featuring a matrix size of 256x256x16. These images were captured at a temporal resolution of 4.8 seconds, consisting of a total of 65 frames recorded over a 5.2-minute timeframe, with the injection of the contrast agent occurring at the 24-second mark.

To analyze this wealth of data effectively, we implemented a sophisticated sliding window approach, skillfully generating 3x3 voxel patches from our DCE-MRI scans of the brain. These patches were meticulously crafted to serve as input for our trained CIF network, which was meticulously designed to yield the Capillary Input Function for the central voxel within each patch.

Following this pivotal step, we embarked on an in-depth voxel-wise pharmacokinetic model analysis. Leveraging the Contrast Transfer Model (CXM), we rigorously analyzed the data, incorporating the voxel-wise CIF denoted as $C_p(t)$, which was thoughtfully provided by our deep neural network. This comprehensive approach encompassed both C_p -TCM and C_p -EVM analyses, enabling us to explore the richness and complexity of the data to the fullest extent.

3 Results

3.1 Loss of estimating CIF



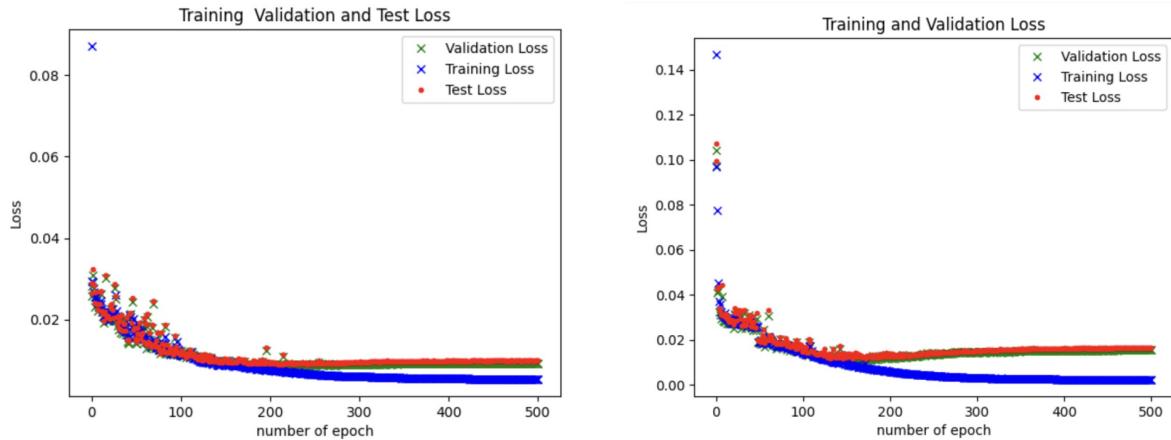


Figure 6: Loss in different training

The decision to utilize only a subset of the complete dataset during training was primarily influenced by time constraints, necessitating a more efficient approach. In this context, we employed two different configurations: (a) where 40,000 data points were used for training, alongside 10,000 for validation, and 5,000 for testing; and (b) a scenario with a reduced training set, encompassing 10,000 data points, while utilizing 2,000 for validation and 1,000 for testing.

	MSE	RMSE	MAE
Train = 10000, Valid = 2000, Test = 1000	0.016307497628538236 ± 0.04755296007627255	0.08995956062861964 ± 0.09063539639701447	0.07641772095781862 ± 0.08244533993220553
Train = 40000, Valid = 10000, Test = 5000	0.009830048773607172 ± 0.02389670884746961	0.0724932971058905 ± 0.06763705085472209	0.06048538089483601 ± 0.06142520859081968

Figure 7: MSE of the Loss

The significance of this choice becomes apparent when examining the impact on the network's performance. In the case of a larger training set (as in scenario (a)), it becomes evident that the results are notably superior. This advantage is marked by smaller losses during training and more robust, stable results during testing and validation. In essence, a larger training set equips the neural network with a broader and more comprehensive understanding of the underlying patterns within the data, leading to enhanced

generalization, lower errors, and ultimately, a more successful model. This underscores the importance of data volume in machine learning, as a larger and more diverse training set can significantly improve the quality of the learned model.

3.2 MSE between the target and the prediction.

The Mean Squared Error (MSE) serves as a fundamental and extensively employed metric across diverse fields, prominently in machine learning and statistics, for evaluating the precision of predictive models. It precisely quantifies the average of the squared disparities between the actual target values and the corresponding predictions generated by the model. In our context of estimating kinetic parameters, this metric plays a crucial role. Here, it is utilized for the essential task of assessing the accuracy of our network's predictions by comparing them with the 'ground-truth CIF' which serves as the target. Then calculate the MSE of PS between predictions and ground-truth

The MSE, in essence, offers an insightful measure of how well a model aligns with the underlying data, with a lower MSE indicating a closer match between predictions and real-world outcomes. Conversely, a higher MSE signifies a less precise model, highlighting substantial deviations between the predictions and the true values. This particularity of the MSE is of paramount significance, making it a staple tool for researchers and practitioners. Especially in the realm of regression tasks, where the aim is to minimize this metric, the MSE serves as a dependable yardstick for evaluating the quality and dependability of predictive models. Its pivotal role is in underlining the need for models to produce accurate and consistent results, promoting their utility and effectiveness.

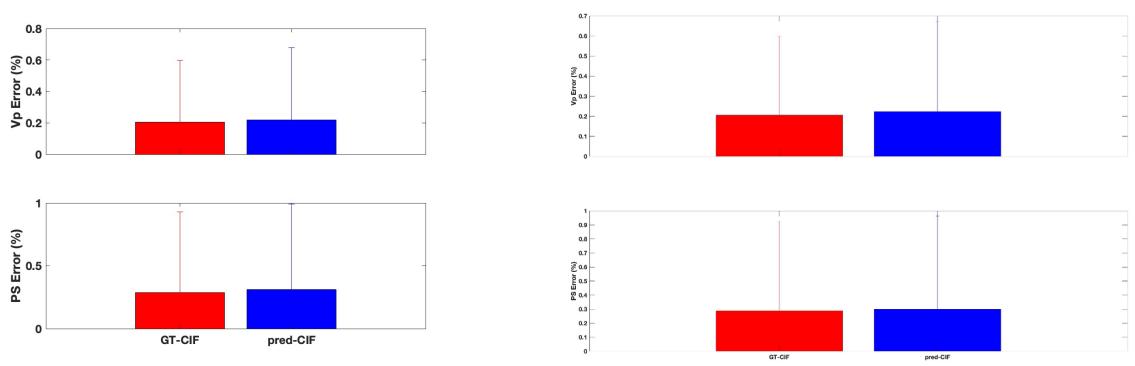


Figure 8: MSE between prediction and ground truth in different trainings

These two figures provide a clear and compelling visual representation of how the Mean Squared Error (MSE) results are notably improved when a larger dataset is used

for analysis. It becomes evident that the expansion in data volume leads to a significant enhancement in the performance of the predictive model.

3.3 Contrast with AIF

To validate the effectiveness of estimating the Capillary Input Function (CIF) through neural network training, we opted to intercept a selection of samples from the test dataset. Our objective was to compare the difference between the Arterial Input Function (AIF) and CIF for the same patients and the same time in different scan. The comparative analysis revealed a striking outcome: the CIF, obtained through neural network training, consistently outperformed the PS values calculated using the AIF for the upper half of the data. The left is the first scan and the right is the second. This outcome was especially pronounced in terms of PS similarity among the same patients. This finding not only confirmed but also illustrated the significance of utilizing neural networks for CIF estimation, particularly in mitigating the two issues highlighted in the introduction. The utilization of CIF, obtained through neural network techniques, effectively reduces the substantial variations in AIF, resulting in more accurate and consistent measurements, a crucial factor in the assessment of dynamic contrast-enhanced MRI data.

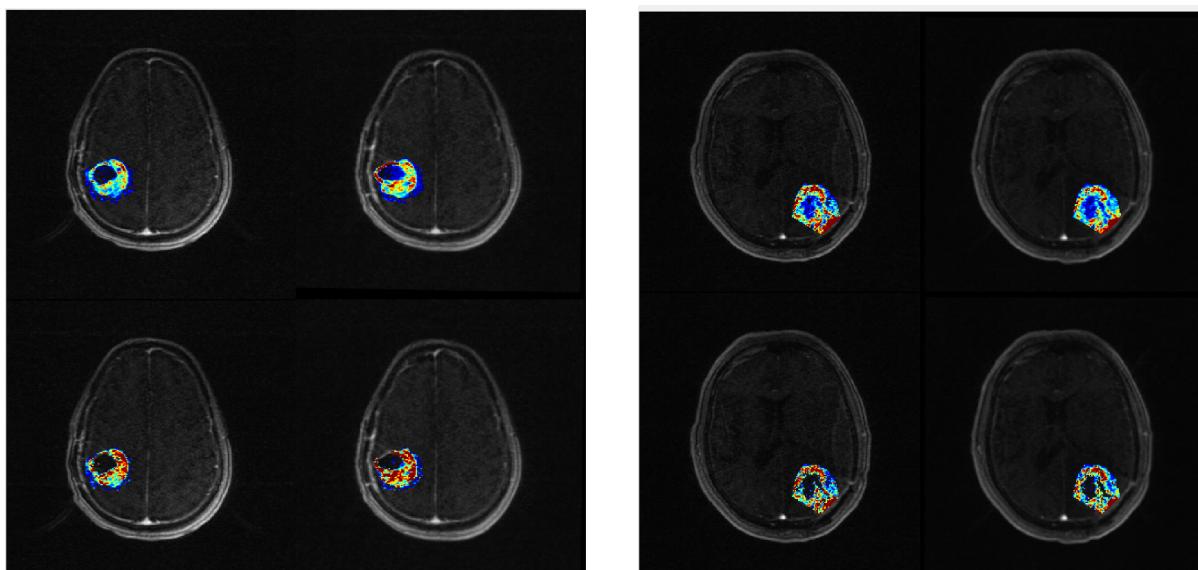


Figure 9: The same patients and the same time in different scan

4 Discussion

Accurate measurement of the Arterial Input Function (AIF) has posed a substantial challenge in the realm of pharmacokinetic model analysis of Dynamic Contrast-Enhanced

MRI (DCE-MRI) data. The AIF typically exhibits the most rapid signal enhancement within a given DCE-MRI dataset and features a significantly higher peak contrast enhancement, which places stringent demands on data acquisition techniques. These requirements involve addressing the temporal resolution, handling the T2* effect, and necessitating a higher spatial resolution to pinpoint voxels within arterial vessels for AIF measurement without partial volume effects. Even in studies that meticulously meet these rigorous acquisition demands, establishing a precise connection between the contrast agent concentration measured at the arterial level (i.e., global AIF) and the concentration at the capillaries of a specific tissue voxel (i.e., local Capillary Input Function - CIF) remains a formidable challenge.

To tackle these considerable hurdles, we have adopted a deep-learning approach aimed at estimating voxel-wise CIF for pharmacokinetic model analysis. This method involves the implementation of two vascular transport models, specifically the Two-Compartment Model (TCM) and the Extended Vascular Tree Model (EVM). In this proof-of-concept study, we have successfully demonstrated the feasibility of training a deep neural network to estimate CIF based on signal changes within a tissue voxel.

An insightful analysis of the test dataset unveiled that pharmacokinetic parameters estimated using CIF lead to more accurate results, even under realistic and noisy conditions, as compared to estimations using case-specific AIFs. Furthermore, we have illustrated that the parameters obtained through the Contrast Transfer Model (CXM) using CIF achieve diagnostic accuracy in Brain tumor detection that is on par with the best set of kinetic parameters relying on AIF. These outcomes underscore the potential for conducting pharmacokinetic model analysis of DCE-MRI data without the need for AIF, thanks to our innovative deep learning approach that allows for the acquisition of voxel-wise CIF. This approach opens up promising avenues for advancing the precision and effectiveness of DCE-MRI analyses, particularly in the context of brain tumor detection and beyond.[3]

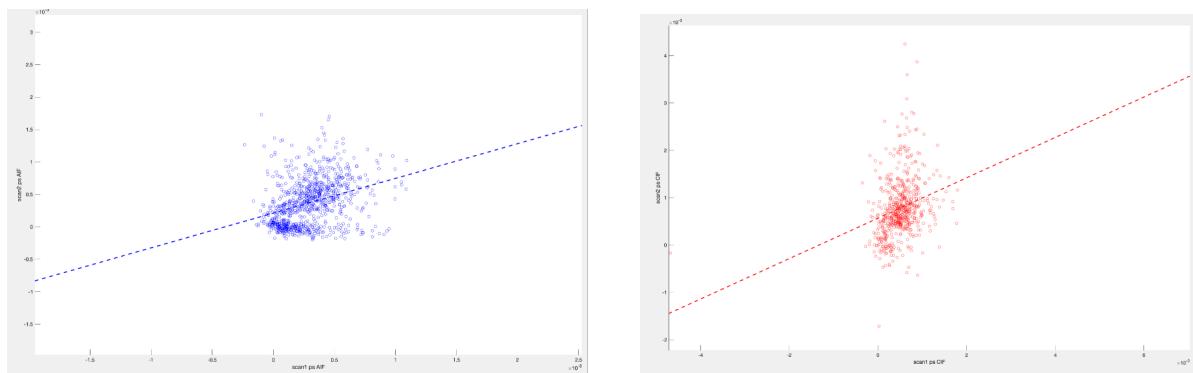


Figure 10: Correlation coefficient of AIF(left) and CIF(left)

However, due to limited training time, the estimated values of some dynamic contrast MRI images trained by neural networks are not as good as AIF. This is because ViT requires a large amount of training data to achieve better performance.

5 Conclusion

The utilization of a Vision Transformer network for estimating voxel Capillary Input Functions (CIFs) has proven to yield superior results compared to manually selected Arterial Input Functions (AIFs) in the majority of cases. However, owing to the constraints of limited training time, there are instances where the results may not reach their full potential. This limitation becomes apparent when examining the covariance of the Permeability-Surface Product (PS) measured from two distinct graphs. It is essential to recognize that for ViT neural networks, the size of the dataset plays a crucial role in the quality of results obtained. In general, a larger dataset often leads to more favorable outcomes, underlining the significance of data volume in enhancing the performance of these neural networks.

References

- [1] Carmel Hayes, Anwar R Padhani, and Martin O Leach. Assessing changes in tumour vascular function using dynamic contrast-enhanced magnetic resonance imaging. *NMR Biomed.*, 15(2):154–163, April 2002.
- [2] Wei Huang, Yiyi Chen, Andriy Fedorov, Xia Li, Guido H Jajamovich, Dariya I Malyarenko, Madhava P Aryal, Peter S LaViolette, Matthew J Oborski, Finbarr O’Sullivan, Richard G Abramson, Kourosh Jafari-Khouzani, Aneela Afzal, Alina Tudorica, Brendan Moloney, Sandeep N Gupta, Cecilia Besa, Jayashree Kalpathy-Cramer, James M Mountz, Charles M Laymon, Mark Muzi, Paul E Kinahan, Kathleen Schmainda, Yue Cao, Thomas L Chenevert, Bachir Taouli, Thomas E Yankelev, Fiona Fennessy, and Xin Li. The impact of arterial input function determination variations on prostate dynamic contrast-enhanced magnetic resonance imaging pharmacokinetic modeling: A multicenter data analysis challenge. *Tomography*, 2(1):56–66, March 2016.
- [3] Jonghyun Bae, Zhengnan Huang, Florian Knoll, Krzysztof Geras, Terlika Pandit Sood, Li Feng, Laura Heacock, Linda Moy, and Sungheon Gene Kim. Estimation of the capillary level input function for dynamic contrast-enhanced MRI of the breast using a deep learning approach. *Magn. Reson. Med.*, 87(5):2536–2550, May 2022.
- [4] F. Ziayee, A. Müller-Lutz, J. Gross, M. Quentin, T. Ullrich, P. Heusch, C. Arsov, R. Rabenalt, P. Albers, G. Antoch, H.J. Wittsack, and L. Schimmöller. Influence of arterial input function (aif) on quantitative prostate dynamic contrast-enhanced (dce) mri and zonal prostate anatomy. *Magnetic Resonance Imaging*, 53:28–33, 2018.
URL: <https://www.sciencedirect.com/science/article/pii/S0730725X18302121>,
doi:10.1016/j.mri.2018.06.004.
- [5] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P Recht, Daniel K Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magn. Reson. Med.*, 79:3055–3071, 2018.
doi:10.1002/mrm.26977.