DISSERTATION

Bayesian Framework for Aleatoric and Epistemic Uncertainty Quantification of Quantitative Parameters: Application to DCE-MRI of the Liver

Bayes'scher Ansatz zur Quantifizierung der aleatorischen und epistemischen Unsicherheit von quantitativen Parametern: Anwendung bei der DCE-MRT der Leber

zur Erlangung des akademischen Grades
Doctor of Philosophy (PhD)

vorgelegt der Medizinischen Fakultät
Charité – Universitätsmedizin Berlin

von

Edengenet Mashilla Dejene

Erstbetreuung: Prof. Dr. med. Winfried Brenner

Datum der Promotion: 28.02.2025

Table of contents

List of tables

List of figures

List of abbreviations

ADAM	Adaptive moment estimation
AI	Artificial intelligence
AIC	Akaike information criterion
AIF	Arterial input function
ANN	Artificial Neural Network
BAT	Bolus Arrival Time
BNN	Bayesian Neural Network
CA	Contrast Agent
CIFAR	Canadian Institute for Advanced Research
CT	Computed Tomography
TCC	Time-concentration curve
DCE	Dynamic contrast-enhanced
DL	Deep Learning
EES	Extracellular Extravascular Space
eTofts	Extended Tofts
FOV	Field-of-view
GPU	Graphics processing unit
GRPE	Golden radial phase encoding
ID	In-Distribution
KL	Kullback–Leibler
MNIST	Modified National Institute of Standards and Technology
MRI	Magnetic Resonance Imaging
NL	Noise level
NLLS	Non-linear-least-squares
OD	Out-of-distribution
PET	Positron emission tomography
RMSE	Root-mean-squared-error
RTX	Ray tracing (graphics)
SENSE	SENSitivity Encoding
TE	Echo time
TR	Repetition Time
XAI	Explainable artificial intelligence

Abstract

Background. Advances in medical imaging play a major role in accurate diagnosis of cancer and treatment planning. Dynamic contrast enhanced (DCE)-MR imaging for example can be used to quantitatively analyse tumor and vascular structures in clinical oncology. The recent advances in Deep-Learning (DL) techniques had shown great progress in quantitative parameter estimation. However, quantitative analysis of DCE-MR images is challenged by the lack of high-quality training data, the need of techniques robust to noise in the data, and difficulty in interpretation of the quality of the obtained quantitative parameters. In the DL-process, the intrinsic ambiguity of the data propagates to the estimated quantitative parameters. Moreover, the training data of the DL-network may mismatch the application data. These factors lead to uncertain parameter estimates.
Aim. The aim of this work is to develop a DL-framework for a joint estimation of quantitative parameters and their respective uncertainties.
Methods. The work was composed of three tasks: i) Simulation of time-concentration curves by applying quantitative parameters and arterial input functions extracted from in-vivo data to a tracer-kinetic model (i.e., extended Tofts (eTofts) model); ii) Training of Bayesian Neural network (BNN) using the simulated data; and iii) Inference of the parameters and uncertainties by application of in-vivo DCE-MR images. The images were acquired for five male patients (56 ± 8 years and, 88 ± 11kg) with hepatic metastases. The estimated quantitative parameters by the proposed framework were compared with the reference method (i.e., non-linear-least-squares fit (NLLS fit)).
Results.The BNN provided more accurate quantitative parameter estimates (for simulated data) and provided parameter estimates more robust to noise of the input concentration curves (for in-vivo data) as compared to the NLLS fitting. The RMSE for the BNN were smaller than the NLLS fit by $33\% \pm 1.9\%$, $22\% \pm 6\%$, $89\% \pm 5\%$ for k_{trans}, v_e and v_p on overage for 0% to 15% noise levels, respectively. The AIC values for the BNN were smaller by $28.5\% \pm 7.5\%$, $12.7\% \pm 6\%$, $12.2\% \pm 2.7\%$ than the NLLS fit for k_{trans}, v_e and v_p on overage for 0% to 15% noise levels, respectively. The quantitative parameters yielded increased aleatoric uncertainties when the noise level in the time-concentration curves was increased. The epistemic uncertainty increased when there was a mismatch between the training and application data or less training data.

Conclusion. The proposed framework provided more accurate quantitative parameter estimates than the NLLS fit, and uncertainty estimates, which explained the intrinsic ambiguity of the data (aleatoric uncertainty) and inadequacy of the trained DL-network to characterize the in-vivo data (epistemic uncertainty).

Zusammenfassung

Hintergrund: Fortschritte in der medizinischen Bildgebung spielen eine wichtige Rolle in der genauen Diagnose von Krebs und in der Behandlungsplanung. So kann die dynamische kontrastmittelbasierte MRT (DCE-MRT) beispielsweise zur quantitativen Analyse von Tumor- und Gefäßstrukturen in der klinischen Onkologie eingesetzt werden. Jüngste Fortschritte bei den Deep Learning (DL)-Methoden haben zu großen Fortschritten bei der quantitativen Parameterschätzung geführt. Die quantitative Analyse von DCE-MRT-Aufnahmen wird jedoch durch den Mangel an qualitativ hochwertigen Trainingsdaten, die Notwendigkeit von Techniken, die gegenüber Rauschen in Daten robust sind, und die Schwierigkeiten bei der Interpretation der Qualität der erhaltenen quantitativen Parameter erschwert. Bei der DL-Methode überträgt sich die inhärente Mehrdeutigkeit der Daten auf die geschätzten quantitativen Parameter. Außerdem kann es vorkommen, dass die Trainingsdaten nicht mit den Anwendungsdaten übereinstimmen. Diese Faktoren führen zu unsicheren Parameterschätzungen.

Ziel: Ziel dieser Arbeit ist es, ein DL-Framework für die gemeinsame Schätzung quantitativer Parameter und deren Unsicherheiten zu entwickeln.

Verfahren: Die Arbeit bestand aus drei Aufgaben: i) Simulation von Zeit-Konzentrations-Kurven, indem quantitative Parameter und arterielle Input-Funktionen, die aus in-vivo-Daten extrahiert wurden, auf ein Tracer-kinetisches Modell (d. h. ein erweitertes Tofts (eTofts)-Modell) angewendet wurden; ii) Training eines Bayes'schen Neuronalen Netzwerks (BNN) unter Verwendung der simulierten Daten; und iii) Inferenz der Parameter und Unsicherheiten mithilfe von in-vivo DCE-MRT-Aufnahmen. Die Aufnahmen wurden von fünf männlichen Patienten (56 ± 8 Jahre und 88 ± 11 kg) mit Lebermetastasen gemacht. Die mittels des vorgeschlagenen Frameworks geschätzten quantitativen Parameter wurden mit der Referenzmethode (d. h. Anpassung durch nichtlineare kleinste Quadrate (NLLS-Fit)) verglichen.

Ergebnisse: Das binäre neuronale Netzwerk lieferte genauere quantitative Parameterschätzungen (für simulierte Daten) und robustere Parameterschätzungen in Bezug auf das Rauschen der Input-Konzentrationskurven (für In-vivo-Daten) im Vergleich zum NLLS-Fit. Die mittlere quadratische Abweichung (RMSE) für das BNN war um $33\,\% \pm 1{,}9\,\%$, $22\,\% \pm 6\,\%$, $89\,\% \pm 5\,\%$ kleiner als das NLLS-Fit durchschnittlich für 0 % bis 15 % Rauschpegel jeweils klei ner als die das NLLS-Fit für k_{trans}, v_e und v_p. Die AIC-Werte für das BNN waren um $28{,}5\,\% \pm 7{,}5\,\%$, $12{,}7\,\% \pm 6\,\%$, $12{,}2\,\% \pm 2{,}7\,\%$ kleiner als das NLLS-Fit für k_{trans}, v_e und v_p durchschnittlich für 0 % bis 15 % Rauschpegel. Die aleatorische Unsicherheit der quantitativen Parameter nahm mit steigendem Rauschpegel zu, während die epistemische Unsicherheit zunahm, wenn es eine Diskrepanz zwischen Trainings- und Anwendungsdaten gab oder weniger Trainingsdaten vorhanden waren.

Fazit: Das vorgeschlagene Framework lieferte genauere quantitative Parameterschätzungen als das NLLS-Fit sowie Unsicherheitsschätzungen, die eine Erklärung für die inhärente Mehrdeutigkeit der Daten (aleatorische Unsicherheit) und die Unzulänglichkeit des trainierten DL-Netzwerks zur Charakterisierung der in vivo-Daten (epistemische Unsicherheit) lieferten.

1 Introduction

Cancer, which is the second leading cause of death, is a global health problem with varying distribution of incidence and death rates around the world [1]. The liver is a common site for metastasis of solid cancers such as cancers of the breast, prostate, and sarcomas, because of its physiology which provides tumor cells a favourable environment for growth [2]. For example, the survival analyses conducted by Zhao et al. [3] indicated that the median overall survival of patients with breast cancer liver metastasis was 2 to 3 years. Accurate diagnosis of cancer is essential for providing prognostic information, which has implications for choosing an appropriate treatment plan [4]. The technological advancements in medical imaging modalities such as magnetic resonance imaging (MRI), computed tomography (CT) and positron emission tomography (PET), allow for the diagnosis of tumors and treatment planning without the need of invasive procedures such as biopsy.

Dynamic contrast enhanced (DCE)-magnetic resonance imaging can be used for qualitative, semi-quantitative, and quantitative analysis, e.g. of the liver, to provide information for cancer diagnosis or to monitor therapy response. DCE-MR imaging requires the acquisition of T1-weighted images before, during and after the administration of a bolus of contrast agent (CA). The injected CA changes the MR signal intensity for different tissues differently depending on its local concentration by reducing the tissue T1 relaxation times. Different tissues such as tumor and healthy tissues, respond to the arrival of the CA with different enhancement patterns primarily depending on blood flow. For the qualitative assessment, the signal enhancement curves are visually assessed according to the temporal enhancement patterns. Qualitative analysis is easy to implement, but the visual assessment is subjective and not standardized (e.g. because of different quality of the signal enhancement curves among image acquisition, which depend on factors such as the quality of the CA injection) [5]. Semi-quantitative evaluation techniques compute semi-quantitative parameters, such as initial area under the curve, time to peak, and the slope of the washout curve, to assess cancer progression and monitor therapy response [6]. These techniques are simple to implement and use. However, the physiological information obtained by the semi-quantitative parameters is not specific and suffers from variability [6].

Quantitative analysis techniques analyse contrast agent time-concentration curves, which are extracted from each voxel of the T1-weighted images, with tracer-kinetic models to derive quantitative parameters, such as perfusion rate, vessel permeability, fractional volume of the plasma and tissue [7], [8]. The derived quantitative parameters can be used for quantitative analysis of tumor and vascular structures for diagnosis or to monitor therapy response in clinical oncology [6], [9]. These techniques are not subjective, and provide more accurate physiological information than semi-quantitative techniques [5].

Nevertheless, the acquisition of DCE-MR images in organs such as the liver is challenging especially because of artifacts emerging from respiratory motion. Motion induces misalignment between DCE time frames, which in turn affects the accurate and reliable estimation of quantitative parameters [10]. To overcome motion-related artifacts, image acquisitions are usually performed during breath hold, using respiratory gating, or free breathing techniques are applied [10]. Breath holding is challenging because some patients are unable to hold their breath long enough for the image acquisition period. Respiratory gating approaches are not appropriate for DCE-MR images because they do not fully record the DCE uptake curve [11]. Free breathing techniques such as Golden-angle RAdial Sparse Parallel (GRASP) have enabled continuous free-breathing data acquisition. However, GRASP can be prone to motion blurring especially during deep breathing [10]. These factors limit the acquisition of high-quality data, and, thus, accurate and reliable estimation of quantitative parameters.

Traditionally, tracer kinetic models are fitted to the time-concentration curves (TCCs) to extract quantitative parameters. The traditional non-linear-least-squares (NLLS) fitting methods are dependent on parameter initializations [12], can converge to a local minima [13], are prone to biased parameter estimates [8], are not robust to noise-affected TCCs [12], [13] and are computationally slow [7]. More recently, Deep Learning (DL) networks have been proposed to overcome the limitations of the NLLS fit due to its increased accuracy, generalization ability and computational speed [7], [8], [12], [13], [14]. Deep Learning based methods on the other hand are challenged by the lack of high-quality ground truth data for training of the neural network. Especially in the liver, the large field-of-view (FOV) and respiratory motion can impair the acquisition of high-quality data. To generate high quality training data, Zou et al. [14] proposed to train a DL-network based on simulated data, while Fang et al. [7], Ottens et al. [8] and Herten et al. [13] proposed to combine simulated and in-vivo data for DL-network training. However, the application to in-vivo data is still affected by noise, residual under-sampling artifacts, or imperfection of coil sensitivities.

The accurate estimation of quantitative parameters from DCE-MRI with DL-networks depends on the inherent ambiguity of the data. Noise-affected input TCCs, which enter the DL-network data analysis process may yield inaccurate quantitative perfusion information. Hence, definite diagnosis cannot be obtained from the estimated quantitative parameters alone in some cases, especially when the TCCs are noise-affected, or the in-vivo data differ from the training data of the DL-network. In such cases, quantification of the propagated aleatoric uncertainty (i.e. uncertainty in the estimated quantitative parameters arising from the intrinsic ambiguity of the TCCs) of the quantitative parameters is crucial for in vivo images like the liver where for example, the residual under-sampling artifact is substantial.

The accurate estimation of quantitative parameters from DCE-MRI also depends on the consistency of training and application data. Quantitative analysis of DCE-MR images with tracer-kinetic models requires the knowledge of patient-specific CA concentration in the plasma, i.e., arterial input function (AIF). Deep Learning networks for parameter estimation usually utilize AIFs extracted from several patients [7], [12], [13] or population-based AIFs [8], [15] as a training data. Arterial input functions extracted from several patients do not describe the possible variations in AIFs among new patients. Arterial input function estimation techniques which assume an averaged AIF for all patients suffer from inter- and intra-subject variations in AIF leading to large potentially errors in estimation of quantitative parameters [6]. Studies by Huang et al. [16] observed intra-subject variation of quantitative parameters, e.g. a coefficient of variation of 0.74 for the rate of CA transfer between plasma and interstitial tissue (i.e. k_{trans}) for unadjusted AIFs. Using an AIF from one cohort of patients and applying it to a different cohort does not account for alterations of the AIF caused by for example, the partial volume effect and in-flow effect [16]. Kim et al. [17] computed the root-mean-square error (RMSE) of each individual AIF from a population-based AIF over time. The results showed a RMSE of 0.88 ± 0.48 mM of the individual AIFs for 18 prostate cancer patients. When out-of-distribution (OD) AIFs, which are AIFs different from the training data, are applied during inference by DL-networks, it can lead to inaccuracies in the estimated quantitative parameters [15]. Other than the AIF, in-vivo application data may have noise levels (NL), TCC distributions or other intrinsic characteristics of the data different from the training data of the DL-network. Noise levels in DCE-MR images, for example show variations among patients depending on the patient size, injected CA dose or imaging artifacts [18], [19], [20]. In other cases, the distribution of the training data does not match the distribution of the application in-vivo data. In such cases, the application in-vivo data substantially differs from the training data.

It is challenging to construct a training dataset that ensures the coverage of all possible NLs, AIFs, or TCC distributions for in-vivo data. Out-of-distribution application data, where the training data is different from the in-vivo data may yield inaccurate quantitative perfusion information. Hence, the uncertainty that the trained DL-network is limited to describe the NLs, TCCs or AIFs of in-vivo data, i.e. epistemic uncertainty, must be accounted for reliable estimation of quantitative parameters. Epistemic uncertainty quantifies uncertainty from applying OD in-vivo data and distinguishes it from other sources of uncertainty, such as the aleatoric uncertainty. In addition, quantification of the epistemic uncertainty helps to guide the selection of training data for example in active learning, by providing a metric on the discrepancy between the training and application data [21].

Bliesener et al. [12] proposed to estimate the quantitative parameters and their aleatoric uncertainties by computing the standard deviations of the posterior distributions of the quantitative parameters for each voxel. Uncertainty was underestimated for low noise levels and overestimated for medium and high noise levels, respectively. In addition, the propagation of aleatoric uncertainty from the TCCs to each of the estimated quantitative parameters and the impact on physiological information was not investigated. Epistemic uncertainty, which arises from a lack of knowledge about the application data leading to a discrepancy between training and application data, was not investigated. As to our knowledge, quantification of epistemic uncertainty for DCE-MR physiological parameter estimation has not been investigated before.

In this work, we implemented a unified Bayesian framework for quantifying both the aleatoric and epistemic uncertainty. The aim is to investigate how noise and OD data influence the quantitative parameter estimates. We quantify the uncertainty of estimated quantitative parameters when OD-AIF, OD-NL, or OD-bolus arrival time (BAT) is applied by modelling the epistemic uncertainty. We address the noise in the TCCs and variations in noise levels by quantifying the aleatoric uncertainty. We showed that quantifying the uncertainties helps explaining the outputs of a DL-network and separating the different sources of uncertainty. This work assesses the proposed framework with simulated data and applies it to in-vivo patient data.

1.1 Contributions

The work in this dissertation strives to overcome the aforementioned challenges for quantitative analysis of DCE-MR data. A novel unified Bayesian neural network framework was developed for quantifying the uncertainty of quantitative parameters.

The main contributions of this work are:

i) A framework to estimate quantitative parameters for liver and liver lesion (patho-) physiology was implemented;

ii) Aleatoric uncertainty was investigated with respect to noise and various levels of noise in the time-concentration curves. The implemented DL-network was sensitive to variations of NLs, and aleatoric uncertainties for all quantitative parameters was estimated properly;

iii) Epistemic uncertainty was investigated by evaluating OD-NLs, OD-AIFs, and OD-BAT delays and less training data size. The implemented DL-network estimated the epistemic uncertainty for OD data and less training data properly; and

iv) Clinical applicability of the proposed framework was investigated by applying in-vivo data to the trained network and estimating quantitative parameters, aleatoric and epistemic uncertainty without the need of ground truth in-vivo quantitative parameter values.

2 Theory

In this chapter, we give a summary of the key concepts used for this work.

2.1 Dynamic Contrast Enhanced (DCE) MR imaging

In DCE-MR imaging, a paramagnetic contrast agent (CA), mostly gadolinium-based, is injected to measure the signal variation of a tissue dynamically. When an exogenous CA is administered to the tissue, the relaxation times of the tissue decreases depending on the distribution of the CA [6]. To acquire DCE-MR images, T1-weighted MRI signal intensity, S(t) is acquired dynamically at different time points: before, during and after injection of the CA for the three spatial dimensions (x, y, z). The temporal and spatial resolutions of the data acquisition have to be selected to capture the dynamics of the CA distribution in time and space.

The S(t) values are converted into time-concentration curves, C(t). The conversion can be performed by assuming a linear or non-linear relation between S(t) and C(t). The aim of the conversion is to quantify the kinetics of the CA in the tissue of interest. Here, a linear conversion method proposed by Medved et al. [22] was used and is also shown in equation 1. This conversion technique was applied to calculate the CA concentrations from the signal intensities for the liver [19], [22].

$$C(t) \approx \frac{S(t) - S(0)}{r.T_{1_{ref}}.S(0)_{ref}} \qquad [1]$$

where $C(t)$ are the CA concentrations in the tissue, $S(0)$ are the precontrast signal intensities, r is the CA relaxivity and $T_{1_{ref}}$ the native preconstrast $T1$ value of a reference tissue and $S(0)_{ref}$ are signal intensities of the reference tissue.

2.2 Compartmental modelling

When a CA is injected (e.g., with a catheter), the CA distributes in the vasculature first and then leaks into extravascular space. Compartmental modelling can be used to model parameters describing the CA movement across different compartments such as the plasma and extravascular space (EES) [6]. One of the most commonly used two compartmental model is the Extended Tofts (eTofts) model.

Extended Tofts (eTofts) Model

Clinically relevant physiological parameters, such as transfer rate of the CA between plasma and EES, fractional volume of the tissue and plasma are modelled by eTofts model as shown in equation 2 [23].

$$C\,(t) = v_p C_p(t) + k_{trans} \int_0^t C_p(t' - \Delta t) e^{-\left(\left(\frac{k_{trans}}{ve}\right)(t-t'-\Delta t)\right)} dt' \qquad [2]$$

where $C(t)$ is the tissue CA concentration change over time, $C_p(t)$ is the plasma CA concentration change over time, k_{trans} (minutes^{-1}) is the transfer rate of the CA between plasma and EES, v_e is the fractional volume of the EES, v_p is the fractional volume of the plasma and Δt is the delay in the arrival of the CA at the tissue.

The eTofts model is illustrated hereunder (Figure 2.1.).

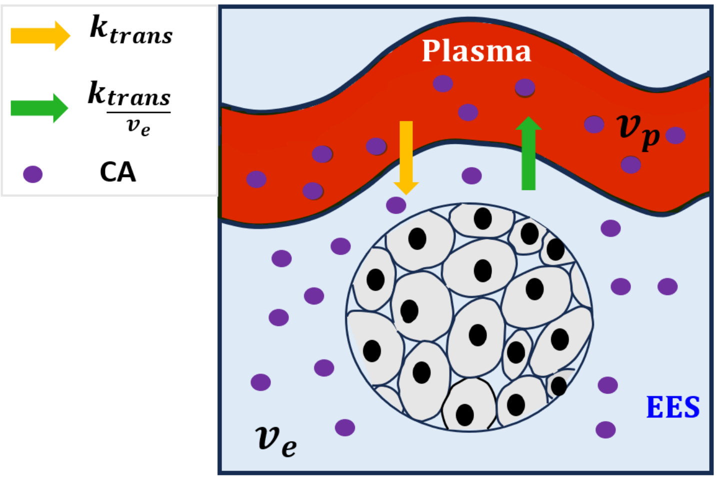


Figure 2.1. Extended Tofts (eTofts) model with plasma and EES compartment. The target quantitative parameters are transfer rate of the CA between plasma and EES (k_{trans}), fractional volume of the EES (v_e) and fractional volume of the plasma (v_p) (own Figure).

Arterial Input Function (AIF)

An accurate estimate of the AIF ($C_p(t)$ in Equation 2), is required to apply compartmental models such as the eTofts model. Arterial Input Function can be estimated by an invasive measurement with a catheter, using a population averaged AIF or obtaining the AIF from the DCE-MR datasets [6]. Measuring the AIF by blood sampling with a catheter enables an accurate measurement of the AIF [6]. However, the procedure is invasive and with a poor temporal resolution. Using a population averaged AIF is simple; there is no need for measuring the AIF for every experimental in-vivo data; and quantitative analysis can be performed on a common AIF for all the available in-vivo data. However, inter- and intra-subject variations in AIF are ignored leading to errors in quantitative analysis. Obtaining the AIF from the DCE-MR images themselves is non-invasive and simple. However, a large vessel within the FOV is required for manual extraction of the AIF from a set of voxels within a vessel.

Bolus Arrival Time delay (BAT)

Bolus Arrival Time Delay refers to the time by which the administered contrast agent arrives to a tissue, such as the plasma or the liver tissue. The BAT between the plasma and liver tissue varies after contrast media injection. For an accurate quantitative parameter estimation, the time delay (Δt) between the bolus arrival in the plasma and liver tissue has to be accounted for (Figure 2.2). The two commonly applied methods [24] to determine Δt are: i) incorporate Δt as a free model parameter in the process of quantitative parameter estimation, and ii) estimate BAT for the plasma and liver tissue separately. The time delay (Δt) is then calculated as the difference between BAT for the plasma and liver tissue. The advantage of the second approach is that it avoids the need of incorporating Δt as a fitting parameter, which would possibly reduce the stability of the fitting. Overall, inaccurate estimation of Δt results in an increased uncertainty in the estimated quantitative parameters [25].

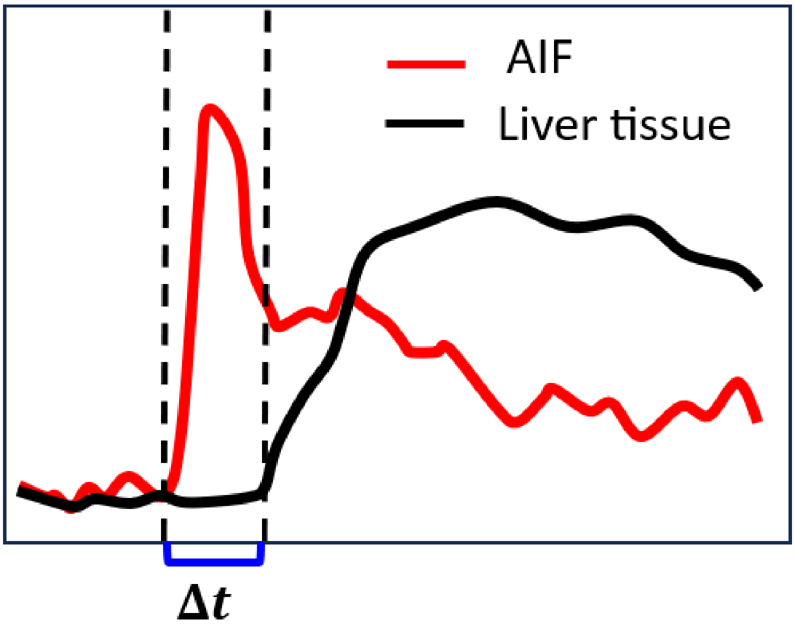


Figure 2.2. *Bolus arrival time delay (Δt) between the AIF and the liver tissue (Own generated Figure).*

2.3 Noise in CA concentration curves

DCE-MR images are affected by noise from various sources. Potential sources of noise include large Field-Of-View (FOV), residual under-sampling artefacts, imperfection of coil sensitivities or acquisition noise [18], [19], [17]. Depending on the patient size and amount of administered CA dose, the noise level varies between patients [18], [19], [17]. When not properly accounted for, the noise in the time-concentration curves can result in inaccurate and imprecise quantitative parameter estimates [9].

2.4 Artificial Neural Networks (ANN)

Artificial neural networks are the basis of DL-networks. An artificial neuron, which is a unit in ANN, functions similarly to the signaling mechanisms of neurons in the brain. A neuron takes a set of inputs $C(t_i)$ with a particular weight w_i and an additional bias b [26]. Based on the sum of the weighted inputs, a neuron applies an activation function f to produce an output z as shown in equation 3 (see Figure 2.3). Inputs with larger weight have more influence on the output z. In the case of DCE-MR analysis, the $C(t_i)$ are the CA concentration values at time point i.

$$z = f\left(\sum_{i}^{n} w_i C(t_i) + b\right) \qquad [3]$$

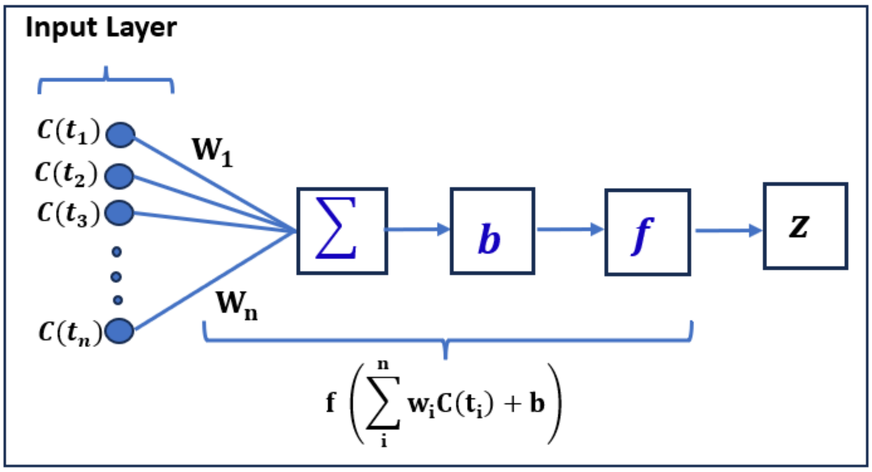


Figure 2.3. A representation of an artificial neuron. Given a set of weighted inputs ($w_iC(t_i)$) and a bias b, the artificial neuron produces the output z (Own generated Figure).

A simple neural network architecture is composed of multiple neurons put together in layers. A DL-network is a neural network consisting of more than two hidden layers. The hidden layers are found between the input and the final output layer. Example of a DL-network with three hidden layers is shown hereunder (Figure 2.4).

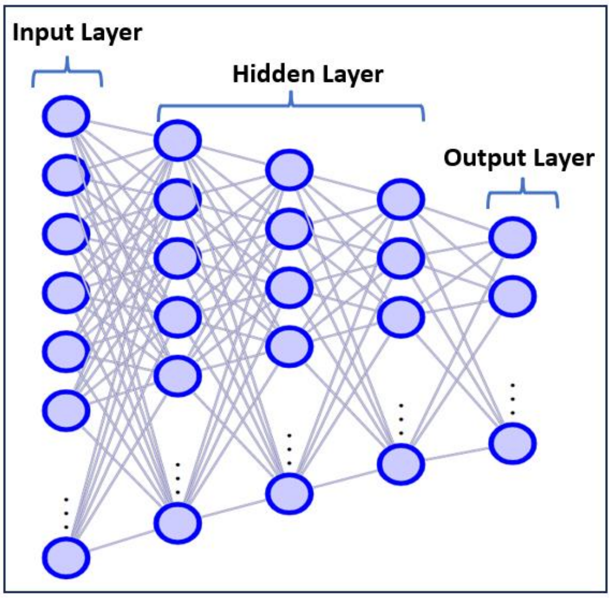


Figure 2.4. A representation of a DL-network with three hidden layers (Own Generated Figure).

Activation Functions

Activation functions are applied to introduce nonlinearity and learn complex representations. Non-linear activations enable to learn non-linear relationships between the input and output neurons [27]. The following functions can be applied as activation functions: sigmoid, rectified-linear unit (ReLU), leaky ReLU, tanh, logistic, softmax, softplus and many more.

2.5 DCE Quantitative Analysis

Quantitative analysis of DCE-MR images is accomplished by fitting the DCE-MRI data or applying DL-networks to extract quantitative information on physiology by using compartmental models. Main steps of quantitative DCE-MR data analysis are shown hereunder (Figure 2.5.).

2.5.1 Conventional Non-Linear-Least-Squares (NLLS) fitting

Non-Linear-Least-Squares fit is the conventional method for tracer-kinetic modelling of DCE-MR data. A tracer-kinetic model, such as the eTofts model is fitted to the ground truth CA concentration curves to extract quantitative parameters. The sum of squared residuals between the ground truth and the predicted time-concentration curves are minimized (as shown in equation 4) for each voxel using optimization algorithms (e.g. NelderMead simplex as proposed by Saravanan et al. [28]). When the algorithm converges, it yields a point estimate of the quantitative parameters ($\hat{\theta}_j$) for each voxel.

$$min \sum_{j=1}^{N} (C_j(t) - \hat{C}_j(t, \hat{\theta}_j))^2 \qquad [4]$$

where $C_j(t)$ are the ground truth concentration curves, $\hat{C}_j\left(t, \hat{\theta}_j\right)$ are the predicted concentration curves and $\hat{\theta}_j$ are the estimated quantitative parameters for voxel j.

2.5.2 Deep Learning (DL) based analysis

Deep Learning-networks provide an alternative approach for quantitative analysis of DCE-MR images. Deep Learning approaches can be distinguished into three main groups, namely supervised, unsupervised and semi-supervised learning [29]. In supervised learning, a labelled data is required for training of the DL-network. Unsupervised learning learns hidden features from the unlabelled data, whereas semi-supervised learning makes use of some labelled training data. The quantitative analysis applied in this work is based on a supervised DL approach.

Supervised Deep Learning (DL)

Supervised DL-networks have recently been used for quantitative parameter estimation [14], [7]. In supervised learning, the ground truth quantitative parameter values are used in the training process of the DL-network to guide the training.

For a training dataset $D = \{(C_j(t), \theta_j)\}$, a DL-network (ϕ) learns a mapping of an input $C_j(t)$ to an output, $\hat{\theta}_j$ where , $\hat{\theta}_j \approx \theta_j$ by using the input-output pairs from the training dataset as shown below:

$$\phi(C_j(t), w) = \hat{\theta}_j \qquad [5]$$

where $C_j(t)$ are the concentration curves, and $\hat{\theta}_j$, are the estimated quantitative parameters for each voxel j and w are the weights of the DL-network. The weights are optimized by minimizing a loss function, which aims to minimize the difference between the ground truth and estimated quantitative parameters. The formulation of the loss function (L) can be written as:

$$L(\theta_j, \hat{\theta}_j) = \min_{\phi} \sum_{j=1}^{N} L(\theta_j, \phi(C_j(t), w)) \qquad [6]$$

The loss function can take different forms. The different losses to be used for DL-network training are discussed in heteroscedastic noise model and Bayesian inference sections of this chapter.

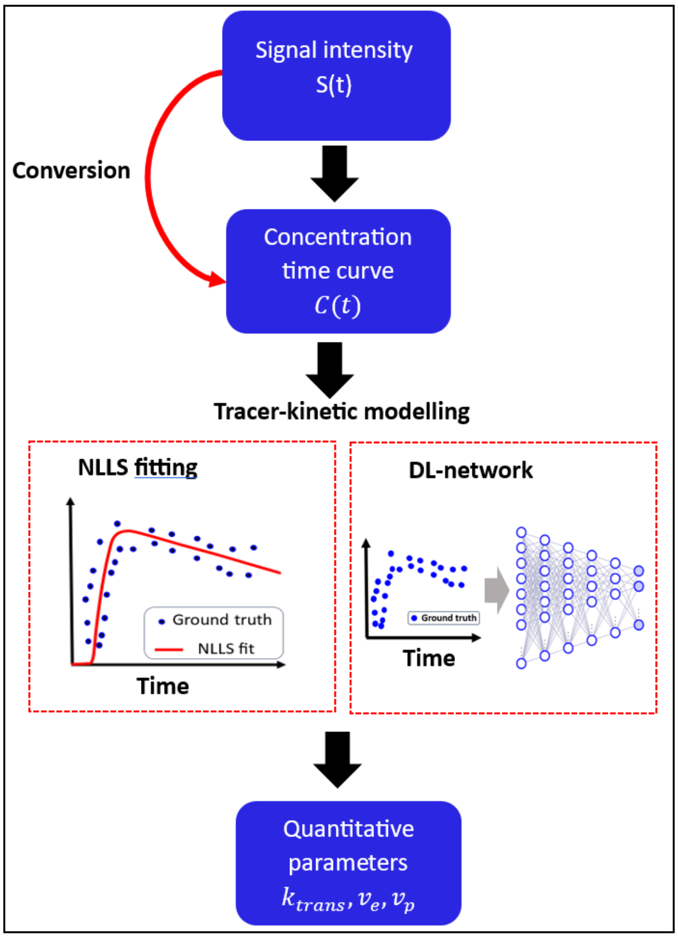


Figure 2.5. Main steps of quantitative DCE-MR data analysis with a NLLS fitting and DL-network (Own Generated Figure).

2.6 Uncertainty quantification

A reliable quantification of the uncertainty is one of the desirable feature of DL-networks, especially for safety-critical applications, such as in medicine [21]. Various sources of uncertainty occur in DL-problems, and depending on the application the handling of the uncertainty also varies. In this work, predictive uncertainty in the context of supervised learning is discussed. The predictive uncertainty is the uncertainty in the estimated quantitative parameters $\hat{\theta}_j$ given by the DL-network from an input data $C_j(t)$. The predictive uncertainty is decomposed into two types of uncertainty, aleatoric and epistemic [21] as shown in Figure 2.6.

Aleatoric uncertainty

Aleatoric uncertainty represents variations in an outcome of an experiment ($\hat{\theta}_j$) arising from the intrinsic ambiguity of the data $C_j(t)$ [21]. Possible sources of this uncertainty are for example, noise in the DCE-MR data [18], [19], [17], which propagates to the estimated quantitative parameters. Even with having a sufficient training data, the aleatoric uncertainty of the estimated quantitative parameter ($\hat{\theta}_j$) can not be reduced.

Epistemic uncertainty

Epistemic uncertainty represents uncertainty arising from lack of knowledge about the application data and is reduced by increasing the size of the training data [21]. Epistemic uncertainty describes cases where the training data is insufficient to accurately represent the data distribution. There are different methods to quantify the epistemic uncertainty such as Gaussian processes, Ensemble methods, Monte Carlo drop out and Bayesian neural networks [30]. In section 2.8., the Bayesian networks, which are applied in this work are discussed.

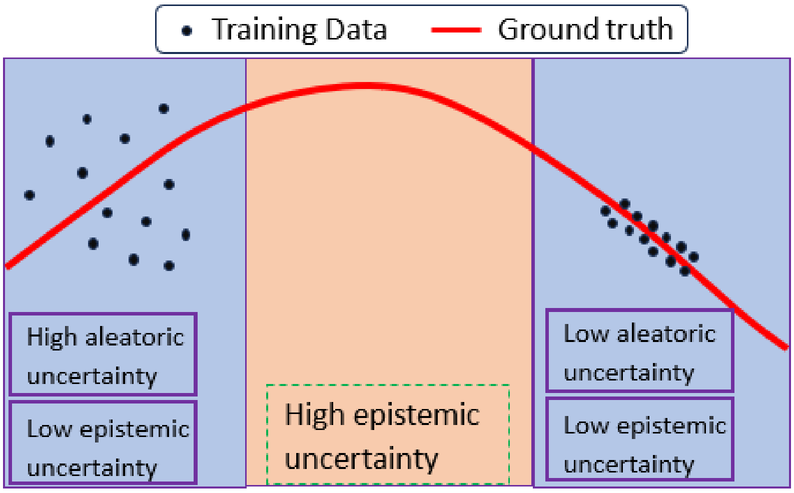


Figure 2.6. Aleatoric and epistemic uncertainty with respect to intrinsic ambiguity and size of training data, respectively (Own GeneratedFigure).

2.7 Heteroscedastic noise model

The heteroscedastic noise model takes into account non-constant noise variance in space and time [31]. The noise variance (σ_{a_j}) is the parameter that has to be estimated for the quantitative parameter, θ_j for each voxel j to get an accurate measure of the noise. It can be predicted as an output of a DL-network. Several approaches have been proposed in the machine learning community for modeling heteroscedastic noise. Here, we show the noise model based on Gaussian likelihood function [31]. The heteroscedastic noise model, shown in equation 7 can be used as a loss function for training a DL-network. When the error between the reference parameter θ_j and estimated parameter $\hat{\theta}_j$ increases, the impact of the uncertainty σ_{a_j} becomes large as shown in the first term. The $\log\left(\sigma_{a_j}\right)$ controls σ_{a_j} from becoming infinitely large, whereas the $\log(2\pi)$ denotes a constant from the Gaussian likelihood function [31].

$$L = \min_{\phi} \frac{1}{2}\sum_{j=1}^{n}\left(\frac{\theta_j-\hat{\theta}_j}{\sigma_{a_j}}\right)^2 + \sum_{j=1}^{n} log\ (\sigma_{a_j}) + \frac{n}{2} log\ (2\pi) \qquad [7]$$

where θ_j are the reference parameters, $\hat{\theta}_j$ are the predicted parameters and σ_{a_j} is the standard deviation of the noise for each voxel j. ϕ is the DL-network with layer weights to be minimized. For the quantitative parameter estimation task, the problem is presented as a voxel-wise estimation of the uncertainty.

2.8 Bayesian Neural Networks (BNN)

Bayesian networks learn a probability distribution for the weights of neural networks.
The core of any Bayesian approach is the computation of a posterior distribution by using the Bayes' rule [32] (equation 8).

$$p\big(w\big|C(t)\big) = \frac{p\big(C(t)\big|w\big)\,p(w)}{p(C(t))} \qquad [8]$$

where p(w) are the priors for the weights of the Bayesian network and p(C(t)) are the probabilities of the data (i.e., concentration curves, C(t)) in tracer kinetic modeling of DCE-MR data. However, the involved numerical integrations for computation of the posterior distribution, $p\big(w\big|C(t)\big)$ are intractable [33]. Variational Bayesian methods apply numerical approximations for the computation of the posterior distribution.
The aim of such variational inference methods is to maximize the agreement between true $p(w)$and approximate posterior distribution $q(w|\beta)$. β is the distribution (e.g., Gaussian with mean and standard deviation) to be learnt during the learning process of the Bayesian training.
The distance between the approximate posterior distribution, $q(w|\beta)$, and the true distribution, $p(w)$, can be measured by applying Kullback–Leibler (KL) divergence loss [33]. The loss function of the Bayesian network training can be written as shown below:

$$L(C(t),\beta) = argmin_{\beta}\, KL[q(w|\beta)\|p(w)] - E_{q(w|\beta)}[\log p(C(t)|w)],$$

which is approximated as:

$$L\big(C_j(t),\beta\big) \approx \sum_{i=1}^{m} log\, q\big(w^i\big|\beta\big) - log\, p\big(w^i\big) - log\, p\big(C_j(t)\big|w^i\big) \qquad [9]$$

where w^i ($w^i \approx q(w^i|\beta)$) are the network weights sampled from the posterior distribution of the network weights to perform a Monte Carlo approximation, $p(C_j(t)|w^i)$ is the likelihood function of the data given the weights. This likelihood function can take different forms (e.g. heteroscedastic noise model) with the aim of minimizing the difference between the ground truth and estimated quantitative parameters.

In this section we discuss Bayesian layers, activation functions, network training procedure and inference which are required for a successful implementation of a Bayesian network.

Bayesian Layers

Bayesian Layers are the building blocks of Bayesian Neural Networks (BNN). The weights of BNN are probability distributions, where each sample from the distribution results in a different weight configuration (see Figure 2.7.) [33]. A Bayesian Layer extends a deterministic layer into a stochastic layer by including a prior distribution over the weights of the network and learning an approximate posterior distribution. These layers capture the uncertainty over the weights. A BNN architecture is composed of multiple neurons stacked in Bayesian layers.

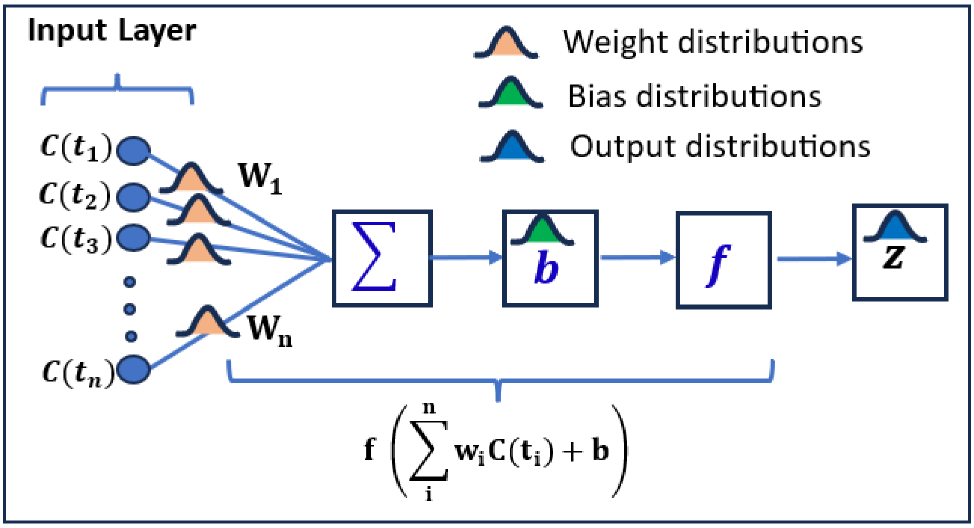


Figure 2.7. A representation of a Bayesian neuron. Given a set of weighted inputs ($w_iC(t_i)$) and a bias b, the Bayesian neuron produces the distribution of output z by computing the sum of the weighted inputs first and then adding the bias (Own Generated Figure).

Network training

To find a minimum for the loss function and parameterize the distribution for the weights, Bayesian networks compute the gradient of the loss in terms of the network weights. This process involves two steps (Figure 2.8): i) forward pass: the input $C_j(t)$ are forward passed to the network to compute the output parameters, $\hat{\theta}_j$ and the loss for this parameter values are computed (L_j); $and\ ii)$ backward pass: gradients of the loss with respect to the probability distribution of the weights, $\beta = (\mu, \alpha)$ are backpropagated (e.g. with stochastic gradient descent or ADAM) from the output to the input layer. These gradients are used to update the variational parameters of the Gaussian distribution, i.e. $\beta = (\mu, \alpha)$, where μ is the mean and α is the standard deviation of the Gaussian distribution β.

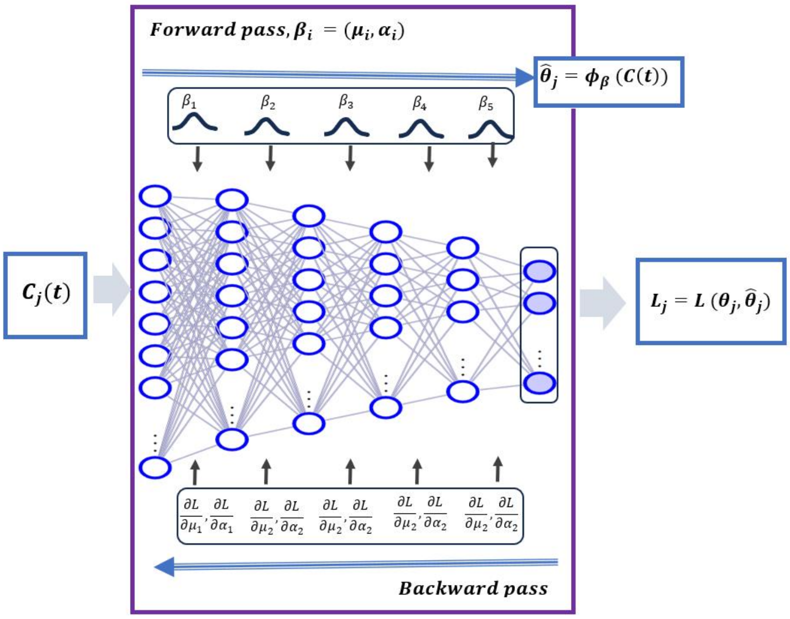


Figure 2.8. *Supervised Bayesian network training with a backpropagation method (Own Generated Figure).*

Inference

The dataset is split into three parts: training, testing and validation dataset. The training dataset is used to learn the parameters of the Bayesian network. The testing dataset is used to evaluate performance of trained network on unseen data. The validation dataset is used to evaluate if the loss values have reached a stable state. During inference, the test dataset is applied to the trained network to infer the outputs as: $\hat{\theta}_j = \phi(C_{j_{test}}(t), \beta)$, where ϕ is the trained BNN, $C_{j_{test}}(t)$ is the concentration curve of the test data for voxel j, β are the probability distributions of the weights of the BNN.

3 Methods

We proposed a unified Bayesian framework, which yields uncertainty information in addition to the quantitative physiological parameter estimates. The proposed framework is shown hereunder (Figure 3.1) (from Dejene et al., 2023 [20]). In this chapter, we discuss how the proposed framework was used for tracer-kinetic analysis and uncertainty quantification.

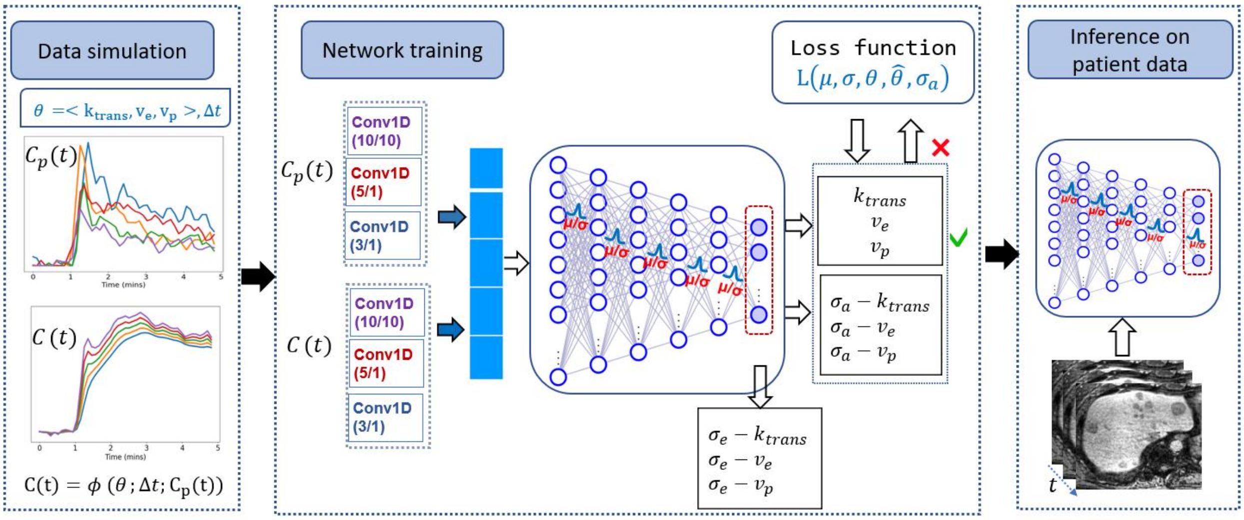


Figure 3.1. The proposed Bayesian framework for aleatoric and epistemic uncertainty quantification of physiological parameters. Figure from Dejene et al., 2023 [20].

3.1 Bayesian Tracer-kinetic Analysis

Time-concentration curves were analyzed to obtain tracer-kinetic information in the form of quantitative parametric maps. The primary goal of applying a BNN was to estimate uncertainties of the quantitative parameter maps for each voxel. To achieve this, the BNN combined $C_p(t)$ and $C(t)$ as input to yield the parameters specifying the CA uptake kinetics, $\theta = \{k_{trans}, v_e\ , v_p\}$, aleatoric $(\sigma_a - k_{trans}, \sigma_a - v_e\ , \sigma_a - v_p)$ and epistemic $(\sigma_e - k_{trans}, \sigma_e - v_e\ , \sigma_e - v_p)$ uncertainties as an output on a voxel level. The BNN architecture is shown above (Figure 3.1., taken from Dejene et al., 2023 [20]). The quantitative parameter maps θ and the aleatoric uncertainties σ_a were the direct outputs of the BNN, whereas the epistemic uncertainty σ_a was computed by executing the trained BNN n times and computing the standard deviation of the n predictions.

The $C_p(t)$ and $C(t)$ were separately convolved by one-dimensional (1D) filters with large, medium and small sizes to extract low, medium and high temporal resolution information, respectively. The filter size and stride length combinations for the large, medium and small sized filters are [10/10], [5/1] and [3/1], respectively. Outputs of the convolutional filters for $C_p(t)$ and $C(t)$ were concatenated and given to the first layer of the BNN as an input. The BNN architecture used 6 fully connected Bayesian layers; each layer consisted of 600, 400, 300, 200 and 100 and 6 neurons. For all the first five layers, Leaky rectified-linear unit (ReLU) activation functions were applied. For the output layer, a sigmoid activation function was applied [20]. This activation function was modified before activation to obtain values between 0 and 2 for k_{trans} to keep outputs in the physiology of the liver [20]. Adam optimizer with 1e^{-5} learning rate was used to train the BNN for 200 epochs. DCE-MRI data was analyzed using python [34]. Bayesian Neural Network was implemented with TensorFlow backend [35] on a Graphics Processing Unit (GPU) workstation (NVIDIA GeForce RTX 2080).

3.2 Bayesian uncertainty quantification

3.2.1 Aleatoric uncertainty

The noise in the TCCs is likely to propagate through the deep learning training process required to estimate the eTofts model parameters. This affects the accuracy and precision of the estimated quantitative parameters [9]. We applied a Gaussian heteroscedastic noise model to handle the aleatoric uncertainty in the estimated quantitative parameters, which were propagated from the noise of the input concentration curves to the estimated parameters. This noise model in equation 7 was applied as a loss function for BNN training. The aleatoric uncertainty $(\sigma_a - k_{trans}, \sigma_a - v_e \text{ and } \sigma_a - v_p)$ for each of the quantitative parameters $(k_{trans}, v_e$ and $v_p)$ is the direct output of the BNN for each voxel j.

3.2.2 Epistemic uncertainty

Epistemic uncertainty occurs if a discrepancy between the training and application data is present [36]. It is the uncertainty arising from the inability of the DL-network to accurately estimate the quantitative parameters because of OD application data (e.g., noise level, AIF or Δt) or less straining data size, ID ($\tau = 0.04$). Bayesian neural networks learn a Gaussian probability distribution, $\beta = (\mu, \alpha)$, for the weights to yield a distribution for the estimated quantitative parameters. The parameter μ is the mean and α is the standard deviation of the Gaussian distribution.

The BNN were repeated n times ($n = 100$) to sample estimates for the physiological parameters from the posterior distribution. The variance was the standard deviation of the predictions when the same Bayesian network was run n different times for the same input. The standard deviations of the BNN estimates for the n iterations were used to quantify the epistemic uncertainty (σ_e) as, $\sigma_{e_k} = \sqrt{\frac{\sum_{j=1}^{n}\left(\hat{\theta}_k - \mu_{\hat{\theta}_k}\right)^2}{n}}$, where σ_{e_k} is the epistemic uncertainty of parameter k, $\hat{\theta}_k$ is the estimated parameter k, $\mu_{\hat{\theta}_k}$ is the mean of the estimated parameter k. The epistemic uncertainty ($\sigma_e - k_{trans}$, $\sigma_e - v_e$ and $\sigma_e - v_p$) for each of the quantitative parameters (k_{trans}, v_e and v_p) was computed during inference by executing the BNN 100 times and computing the voxel-wise standard deviation of the predictions.

3.2.3 Loss function

To jointly estimate the aleatoric and epistemic uncertainties, a combined loss function consisting of the heteroscedastic noise model and KL-divergence loss was applied. The combined loss [20] was written as:

$$L = min\frac{1}{2}\sum_{k=1}^{3}\left(\frac{\theta_k - \hat{\theta}_k}{\sigma_k}\right)^2 + \sum_{k=1}^{3} log\,(\sigma_{a_k}) + \frac{n}{2} log\,(2\pi) + \sum_{k=1}^{3} log\,q\left(w^i \middle| \beta\right) - log\,p\left(w^i\right) \qquad [10]$$

where k is the number of physiological parameters, w^i is the i weight sample taken from the approximate posterior distribution of the weights.

3.3 Experiments

3.3.1 Simulated Data

DCE-MRI data were simulated by applying the eTofts model on physiological parameters, $\theta = \{k_{trans}, v_e$ and $v_p\}$, incorporating BAT delay (Δt) and addition of a Gaussian noise.

Physiological parameters

Time-concentration curves were simulated using parameter combinations of k_{trans}, v_e and v_p in the following ranges, $k_{trans} \in [0,2], v_e \in [0,1]\ and\ v_p \in [0,0.3]$ in step size, τ= 0.02 to confine them within physiological range of the liver [20]. $C_p(t)$ from five patients were manually extracted from the hepatic artery. For a particular combination of parameter values, the $C_p(t)$ from five patients were applied during the TCCs simulation process. A total of 75,000 TCCs were simulated for each $C_p(t)$.

Noise levels (NL)

We added Gaussian noise to mimic the noise distribution in the in-vivo data. Variations of noise levels were added to investigate their effect on the uncertainty. A range of noise levels with mean zero and standard deviation ($\lambda = NL * C(t)_{max}$) [20], were added to the TCCs. NL was the noise level in percentages and $C(t)_{max} = 0.59$ was the maximum value of the concentration curves.

Bolus Arrival Time delay (Δt)

The delay (Δt) between the time of CA enhancement in the plasma and liver tissue was accounted for by shifting the AIF in ranges of three discrete time points (i.e., $\Delta t \in [6s, 12s, 18s]$). These Δt values were applied to simulate $C(t)$ based on equation 2. For a particular combination of parameter values and noise levels, the TCCs were simulated for each Δt values. The shift time, Δt was set upto $18s$ to bound it in the liver physiological BAT delay ranges [37], [38].

The contrast agent arrives earlier at the hepatic artery than the liver tissue [39]. With Δt included in the simulation of TCCs, the time of initial plasma enhancement would shift to a later time point to be matched with the time of the liver tissue enhancement. During training, the BNN learns the temporal shifts in the C(t) due to the difference in the bolus arrival time for different tissues. Hence, Δt was intrinsically incorporated in the learning process and the effect of uncorrected BAT delay on the estimated quantitative parameters was corrected.

3.3.2 In-vivo Data

The DCE-MR in-vivo image acquisition was performed by using a 3T Biograph mMR hybrid scanner (Siemens Healthcare, Erlangen, Germany) [40]. Imaging data for the liver were acquired using a 3D golden-radial phase encoding (GRPE) acquisition by applying a Cartesian sampling scheme [19]. A hepato-specific contrast agent, gadoxeate disodium with a dose of 0.01 mmol kg^{-1} was administered 1 min after image acquisition had begun [19]. “The acquisition parameters were: TR/TE = 3.3 ms/1.36 ms, flip angle = 12°, FOV =

345 X 345 mm^2, spatial resolution = 1.5 mm^2, partial Fourier factor = 5/8.” [20]. Motion corrected image reconstruction based on iterative kt-SENSE [19] was used to reconstruct the DCE-MR images for 47 dynamic scan points. The temporal resolution was 6 seconds. In this work, DCE-MR data from five male patients (56 ± 8 years and, 88 ± 11kg) with hepatic metastases were included [19]. The study was approved by the Charité Ethics Committee and written informed consent was provided by all patients.

3.4 Evaluation on simulated Data

The simulated data was used to train a BNN, investigate accuracy of the proposed uncertainty estimation framework, compare physiological parameter estimates between the NLLS fit and BNN and assess uncertainties of physiological parameter estimates.

3.4.1 Parameter estimation

We compared physiological parameters estimated by the standard NLLS fit and the proposed BNN. For assessing the parameter estimation between the two methods, it was assumed that $\Delta t = 0$.

Non-linear-least-squares (NLLS) fitting

We evaluated the performance of the NLLS fit on different noise levels. NLLS fit optimizes the physiological parameters so that the eTofts model fits best matched the reference $C(t)$. For this, the root-mean-squared (RMSE) between the reference $C(t)$ and predicted $C(t)$ are minimized using the NelderMead Simplex optimization algorithm [28] for each voxel. This results in a single set of parameter values for each TCC in a deterministic setting. Gaussian noise with noise levels ranges of NL ∈ $[0\%, 1\%, 5\%, 10\%, 15\%]$ were added to the TCCs. The initial values of the fitting for k_{trans}, v_e, v_p and Δt were set to $0.8, 0.1, 0.01$ and 0.001, respectively. The lower and upper bounds of the fitting for k_{trans}, v_e, v_p and Δt were set to $[0.001, 0.001, 0.001, 0]$ and $[2, 1, 1, 0.4]$, respectively [20].

BNN estimation

The performance of the BNN was investigated for different noise levels. BNN optimizes the physiological parameters so that the estimated parameters, $\hat{\theta}$ best matched θ. For this, the combined loss function (Equation 10) is minimized. This results in all probable sets of estimated parameter values, $p(\hat{\theta}|C(t))$ instead of providing estimates for a single set of parameter values. Five BNNs are trained with each of Gaussian noise levels, NL $\in [0\%, 1\%, 5\%, 10\%, 15\%]$. The time-concentration curves with a NL similar to the training NL were applied during inference for each BNN.

Evaluation metrics

The performance of NLLS fit and BNN for different NLs were compared by using quantitative metrics. Root-mean-squared-errors (RMSE), R^2 and Akaike information criterion (AIC) values [41] were calculated between estimated physiological parameters (i.e., by BNN and NLLS fit) and reference parameters.

3.4.2 Uncertainty evaluation

We used the simulated DCE-MR data to train the BNN and applied it to test datasets containing ID and OD data.

ID-data ($\boldsymbol{\tau = 0.02}$)

The aim of this experiment was to investigate the change in uncertainty for variations of ID-NLs. In-Distribution (ID) test data are TCCs similar in NL, AIF, Δt and τ to the TCCs applied to the BNN during training. We compared uncertainty estimates at various ID-NLs ($NL \in [0\%, 1\%, 5\%, 10\%, 15\%]$). The training and testing data contain similar NLs ($NL \in [0\%, 1\%, 5\%, 10\%, 15\%]$). τ was set to 0.02. During inference, the BNN with a training data noise level similar to the testing data noise level was applied.

ID-data ($\boldsymbol{\tau = 0.04}$)

The aim of this experiment was to investigate the uncertainty when the training data size was reduced by half. For this, the training TCCs were simulated with the same AIF, NL and Δt as the test data. However, the training data has TCCs generated with τ= 0.04, whereas the test data contain TCCs generated with τ= 0.02.

OD-data

The aim of this experiment was to investigate the uncertainty when a range of OD application data sources (i.e., NL, τ, AIF and Δt) were considered. OD test data are TCCs different either in NL, τ, AIF, or Δt to the TCCs applied to the BNN during training. It was challenging to construct a training dataset that covered all possible NLs, AIF, τ or Δt, especially in the in-vivo setting, where the in-vivo data NLs and TCCs distribution or Δt are unknown. Here, we investigated three OD application data cases, namely, OD-NL, OD-AIF, and OD-Δt and less training data, ID- ($\tau = 0.04$). The training and application data ranges are summarized and tabulated in **Error! Reference source not found.** (Adapted from Dejene et al., 2023 [20]).

Table 3.1. ID and OD data experiments performed for simulated and in-vivo data (modified Table from Dejene et al., 2023 [20])

Data (Train, Test)	Experiments	Train NL (%)	Test NL (%)	Train (step size)	Test (step size)
Simulated, Simulated	ID-($\tau = 0.02$) ID-($\tau = 0.04$) OD-NL OD-AIF OD-Δt	1,5,10,15,20 5,10 5,10 5,10 5,10	1,5,10,15,20 5,10 15,20 5,10 5,10	0.02 0.04 0.02 0.02 0.02	0.02 0.02 0.02 0.02 0.02
Simulated,	ID-($\tau = 0.02$)	10,15,20	–	0.02	–
In vivo	ID-($\tau = 0.04$)	10,15,20	–	0.04	–
	OD-AIF	10,15,20	–	0.02	–
NL – noise levels, ID – In distribution, OD – Out-of-distribution					

i. **OD-NL**

The effect of different NLs on the uncertainties was evaluated. The test data consists of OD-NLs (15% and 20%) different from the training NLs (5% and 10%).

ii. **OD-AIF**

The effect of different AIF (i.e., different amplitude or BAT delays) on the uncertainty was investigated. The test data consists of OD-AIF (peak amplitude broadened by a factor of 2 and Δt =36s) different from the training AIF (no change in amplitude and Δt =0).

iii. **OD-Δt**

The effect of different bolus BAT delays on the uncertainty was investigated. The test data consists of OD-Δt ($\Delta t = 12s$) different from the training Δt ($\Delta t = 0s$).

3.5 Evaluation of in-vivo Data

We evaluated the proposed framework on DCE-MR images of the liver to demonstrate its applicability in in-vivo data.

3.5.1 Parameter estimation

Using in-vivo data, we compared parameter estimation by NLLS fit and BNN estimation.

Non-linear-least-squares (NLLS) fitting

We evaluated the performance of the NLLS fit on in-vivo data. The in-vivo TCCs were fitted to the eTofts model to estimate the quantitative parameters.

BNN estimation

We evaluated the performance of the BNN on in-vivo data. The training data of the BNN consists of data to construct a training dataset that covers reasonably all combinations of NLs, AIF, τ or Δt prevalent in in-vivo data. A BNN with simulated training dataset consisting of $NL \in [10\%, 15\%, 20\%]$, AIFs (taken from five patients), TCCs ($\tau = 0.02$) and $\Delta t = [0s, 6s, 12s, 18s]$ was trained. The in-vivo data was applied to the trained BNN to infer the quantitative parameters.

Time Cost

The execution time of the BNN to yield the quantitative parameters and the uncertainties was evaluated for a DCE-MR scan of an in-vivo data with 192 × 192 X 47 voxels utilizing the Python time module [34] on a GPU. The execution time of the NLLS fit for yielding the quantitative parameters for this same slice was also evaluated on a CPU.

3.5.2 Uncertainty evaluation

We used the simulated DCE-MR data to train the BNN and apply it to in-vivo data consisting of ID and OD data.

ID in-vivo data ($\boldsymbol{\tau = 0.02}$)

The aim of this experiment was to investigate the uncertainty estimates when in-vivo application data is applied to a BNN trained with simulated data. In-Distribution in-vivo data are assumed to be similar in NL, AIF, τ and Δt to the training data of the BNN. The training data of the BNN consists of simulated TCCs with $NL \in [10\%, 15\%, 20\%]$, AIFs (from five patients), $\tau = 0.02$ and $\Delta t \in [0s, 6s, 12s, 18s]$.

ID in-vivo data ($\boldsymbol{\tau = 0.04}$)

The aim of this experiment was to evaluate the effect of decreasing the training data size on the uncertainty of in-vivo data. The training data of the BNN consists of simulated TCCs with $NL \in [10\%, 15\%, 20\%]$, AIFs (from five patients), $\tau = 0.04$ and $\Delta t \in [0s, 6s, 12s, 18s]$. A step size of $\tau = 0.04$ reduces the number of training samples by 50% as compared to the ID training data with a step size of $\tau = 0.02$.

OD in-vivo data

The aim of this experiment was to investigate the uncertainty when OD-AIF in-vivo data are applied to a BNN trained with simulated data. It is not practical to construct a dataset that covers all possible combinations of NLs, AIFs, τ or Δt. Especially the true NLs, τ and Δt in in-vivo data are unknown. OD-in-vivo data are TCCs different in NLs, AIFs, τ or Δt to the training data of the BNN. Since the true Δt and NLs in in-vivo data are unknown, we investigate the uncertainty estimates of OD-AIF only. The training data ranges for these experiments are summarized in tabular form (**Error! Reference source not found.**) (taken from Dejene et al., 2023 [20]). The training data did not constitute the AIF of the application in-vivo data.

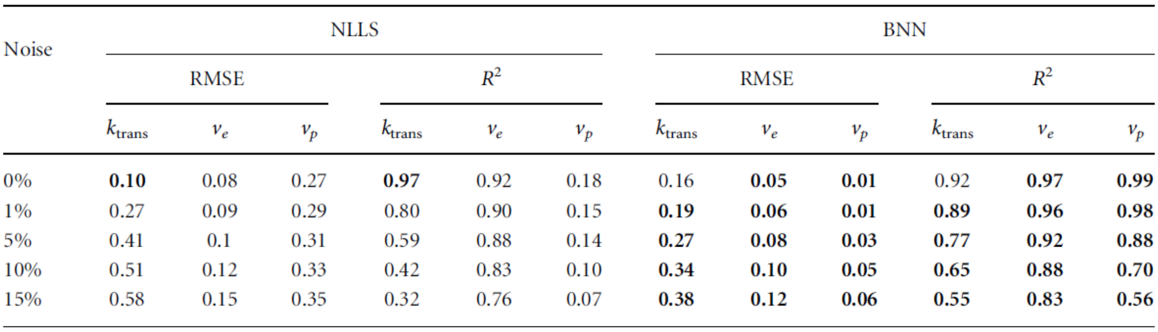
4 Results

4.1 Evaluation on simulated Data

4.1.1 Parameter estimation

The BNN demonstrated better performance than the NLLS fit in terms of R^2, RMSE and AIC values. Dejene et al. [20] presented the comparisons of the RMSE and R^2 values between the reference and estimated parameters for the NLLS fit and BNN for different noise levels (Table 4.1.). The RMSE of k_{trans}, v_e and v_p deteriorated with increasing NLs from 0.27 to 0.58, 0.09 to 0.15 and 0.29 to 0.35 when the NL was increased from 1% to 15% for the NLLS fit. The RMSE of k_{trans}, v_e and v_p deteriorated with increasing NLs from 0.19 to 0.38, 0.06 to 0.12 and 0.01 to 0.06 when the NL was increased from 1% to 15% for the BNN. Overall, the RMSE for the BNN were smaller than the NLLS fit by $33\% \pm 1.9\%$, $22\% \pm 6\%$, $89\% \pm 5\%$ for k_{trans}, v_e and v_p on overage for 0% to 15% noise levels, respectively. The R^2 values for the BNN were higher than the NLLS fit. The AIC values for the BNN were smaller by $28.5\% \pm 7.5\%$, $12.7\% \pm 6\%$, $12.2\% \pm 2.7\%$ than the NLLS fit for k_{trans}, v_e and v_p on overage for 0% to 15% noise levels, respectively.

Table 4.1. Comparison of RMSE and R2 between NLLS fit and BNN. Bold results show the lowest RMSE and highest R2 values (taken from Dejene et al., 2023 [20]).



Noise	NLLS						BNN					
	RMSE			R^2			RMSE			R^2		
	k_{trans}	v_e	v_p	k_{trans}	v_e	v_p	k_{trans}	v_e	v_p	k_{trans}	v_e	v_p
0%	**0.10**	0.08	0.27	**0.97**	0.92	0.18	0.16	**0.05**	**0.01**	0.92	**0.97**	**0.99**
1%	0.27	0.09	0.29	0.80	0.90	0.15	**0.19**	**0.06**	**0.01**	**0.89**	**0.96**	**0.98**
5%	0.41	0.1	0.31	0.59	0.88	0.14	**0.27**	**0.08**	**0.03**	**0.77**	**0.92**	**0.88**
10%	0.51	0.12	0.33	0.42	0.83	0.10	**0.34**	**0.10**	**0.05**	**0.65**	**0.88**	**0.70**
15%	0.58	0.15	0.35	0.32	0.76	0.07	**0.38**	**0.12**	**0.06**	**0.55**	**0.83**	**0.56**

4.1.2 Uncertainty evaluation

ID-data ($\tau = 0.02$)

The aleatoric uncertainty estimates of the quantitative parameters increased with increasing noise levels (Figure 4.1.) (Adapted from Dejene et al., 2023 [20]). The epistemic uncertainty estimates of the parameters were similar between different NLs, showing very small differences on average compared to all NLs (i.e., $0.08 \pm 2.7\%, 0.03 \pm 3.4\%, 0.02 \pm 5\%$ for k_{trans}, v_e and v_p, respectively).

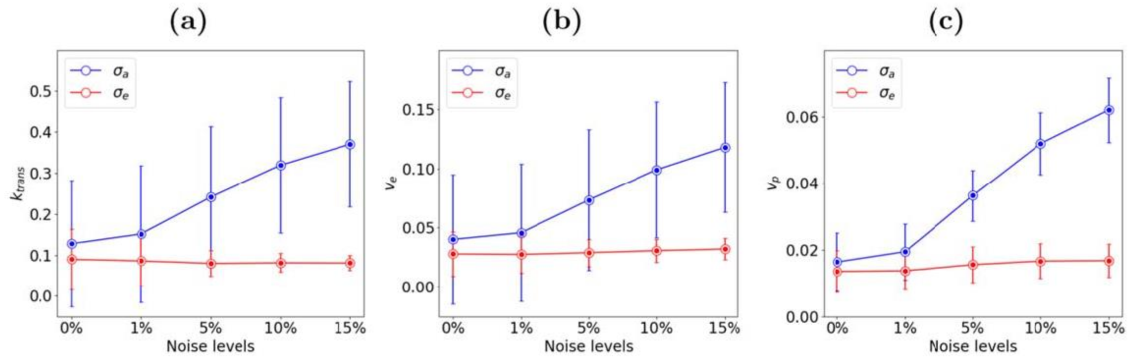


Figure 4.1. Aleatoric (σ_a) and epistemic (σ_e) uncertainty of ID-data for (a) k_{trans} (b) v_e and (c) v_p (taken from Dejene et al., 2023 [20]).

ID-data ($\tau = 0.04$)

The epistemic uncertainty of the quantitative parameters with less training data significantly increased (p<0.001) compared to the ID-data ($\tau = 0.02$) (Figure 4.2., taken from Dejene et al., 2023 [20]).

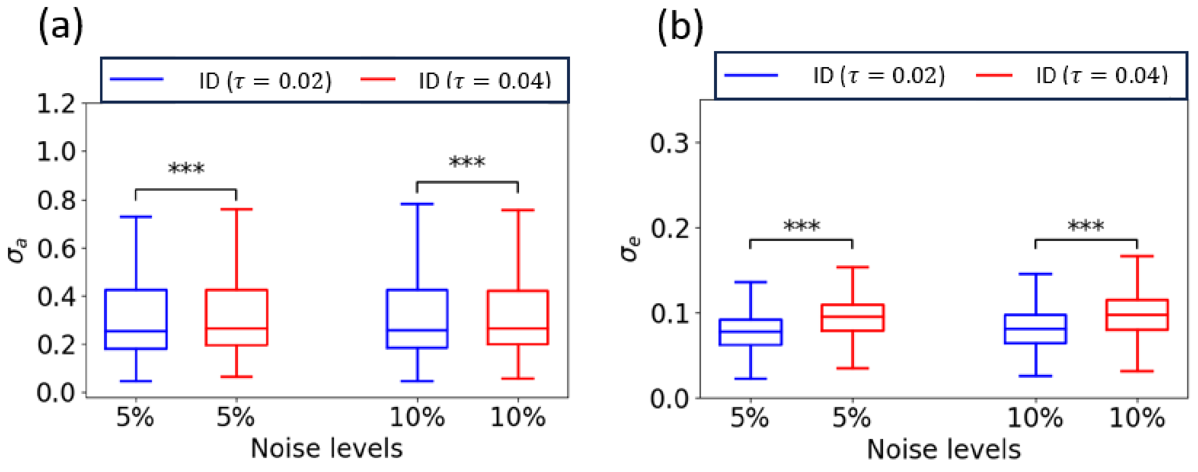


Figure 4.2. k_{trans} aleatoric uncertainty (σ_a) and epistemic (σ_e) for ID-data with less training data size (adapted from Dejene et al., 2023 [20]).

OD-data

The aleatoric and epistemic uncertainty of the quantitative parameters varied because of applying OD-data to the trained BNN. Dejene et al. [20] showed uncertainty evaluation of OD-NL, OD-AIF and OD-Δt (Figure 4.3., (taken from Dejene et al., 2023 [20])). The epistemic uncertainty of the quantitative parameters for the OD-data increased significantly (p<0.001) compared to the ID-data for all the OD cases. Aleatoric uncertainties increased for OD-AIF and OD-Δt and decreased for OD-NLs.

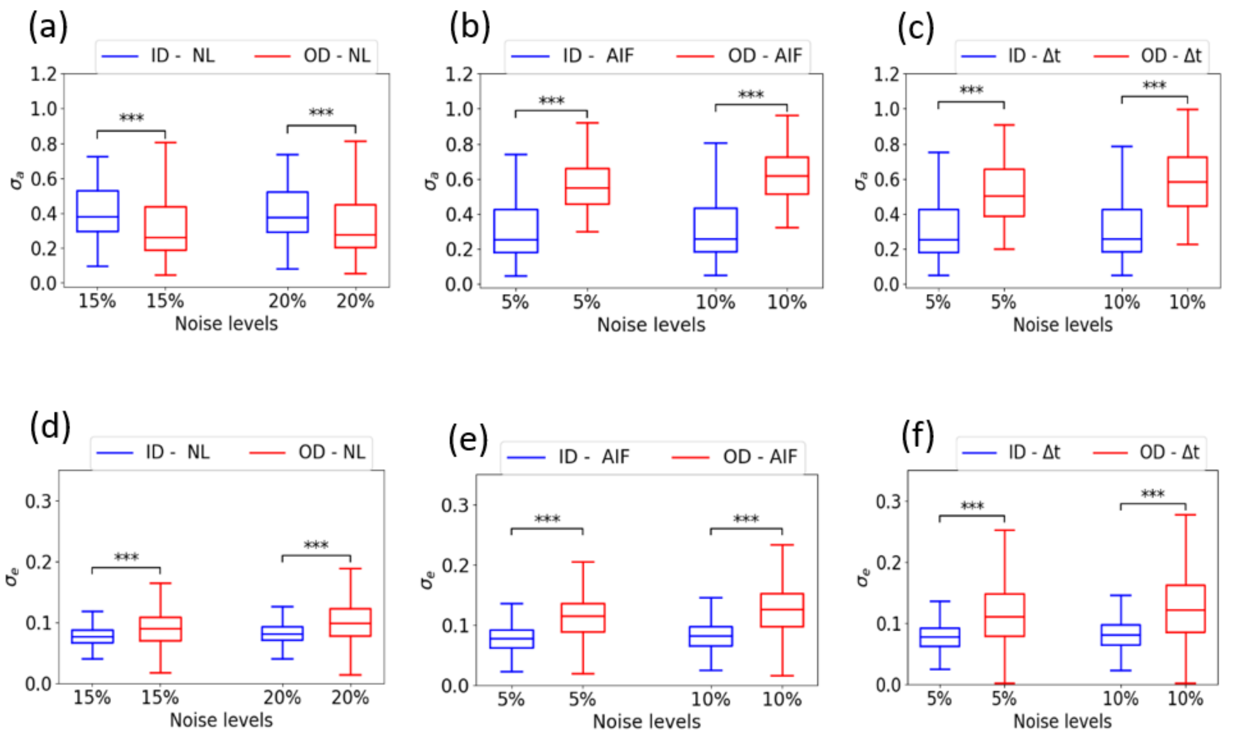


Figure 4.3. k_{trans} aleatoric uncertainty (σ_a) and epistemic (σ_e) uncertainty for OD-NLs (a), (d), OD-AIFs (b), (e), and OD-Δt (c), (f) (taken from Dejene et al., 2023 [20]).

4.2 Evaluation on simulated Data

4.2.1 Parameter estimation

In Figure 4.4. (taken from Dejene et al., 2023 [20]), we presented parameter maps obtained by applying the NLLS-fit and BNN on DCE-MR scans of in-vivo data. Bayesian Neural Network provided parameter estimates more robust to noise in the input concentration curves for k_{trans}, v_e and v_p than the NLLS-fit. The region-of-interest (ROI) assessment for the tumor lesion and the healthy region for the two methods are shown in Figure 4.5. (Adapted from Dejene et al., 2023 [20]). Region-of-interest for tumor lesions yielded high k_{trans} and v_p values and low v_e values in both the NLLS fit and BNN estimation. For

NLLS fit, a large variance of the parameters was observed, especially for k_{trans} and v_e in the tumor lesion and healthy tissue, respectively.

The parameter estimates for less training data and OD-AIF of in-vivo data showed high values for k_{trans} and v_p and low v_e values for the tumour lesions, similar to the results of the ID-data (Figure 4.6. (Adapted from Dejene et al., 2023 [20])). For OD-AIF, overestimation of k_{trans} was observed, while less training data resulted in parameter maps comparable to the ID-data.

Time Cost

The execution time for the BNN to estimate the quantitative parameter maps and the uncertainties was 0.1 min for a DCE-MR scan with $192 \times 192 \times 47$ voxels, whereas the execution time for the NLLS fit to yield the quantitative parameter maps was 270 min. The execution time was considerably improved with the BNN.

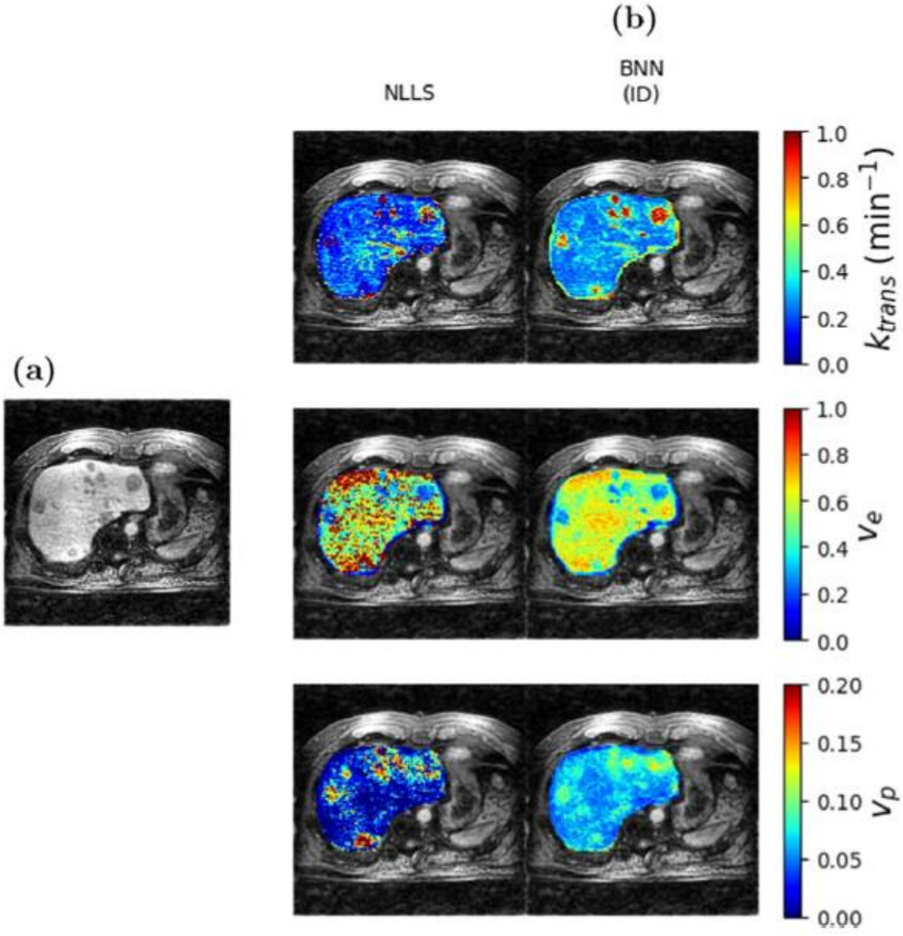


Figure 4.4. (a) DCE-MR scans of the liver with tumor lesions (b) k_{trans}, v_e and v_p estimated by the NLLS fit and BNN (taken from Dejene et al., 2023 [20]) .

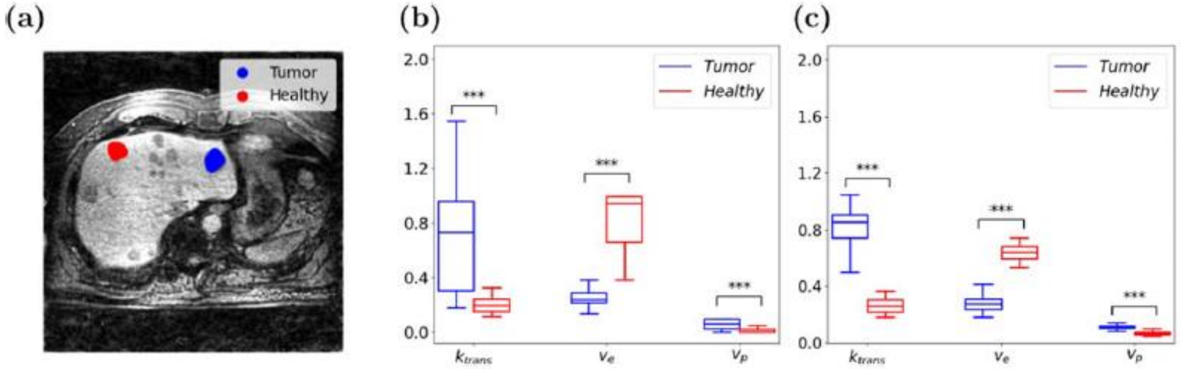


Figure 4.5. ROIs in tumor and healthy regions (a), quantitative parameter estimates (k_{trans}, v_e and v_p) by the NLLS fit (b) and BNN (c) (taken from Dejene et al., 2023 [20]).

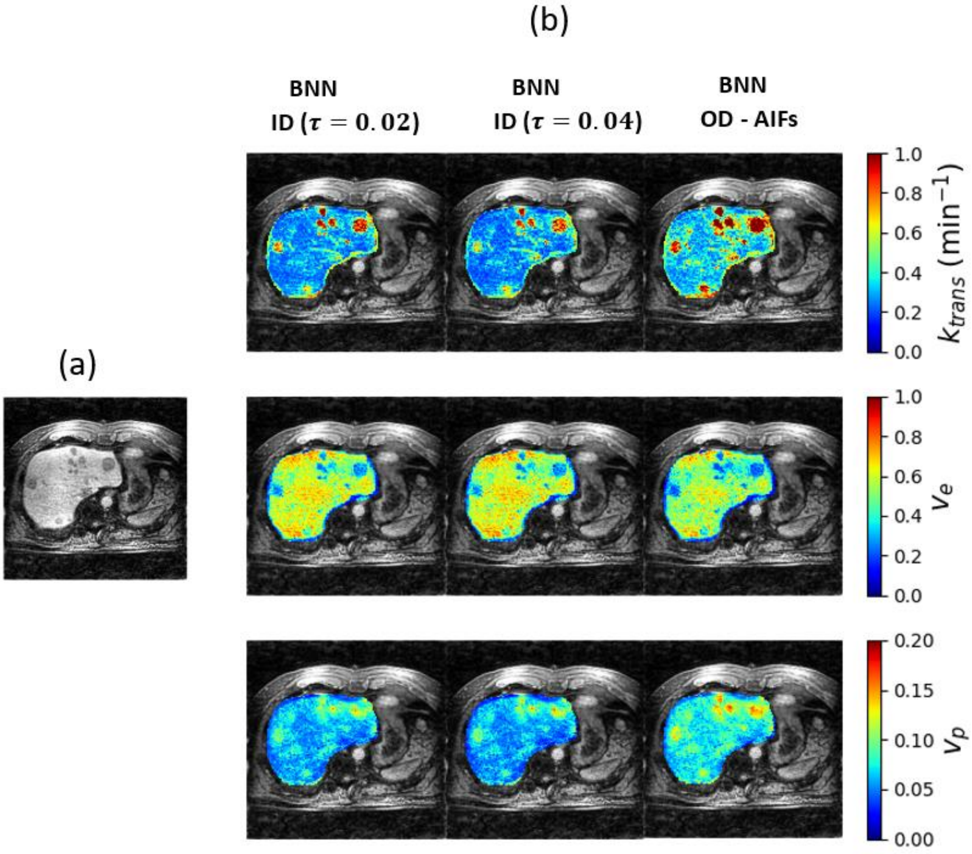


Figure 4.6. (a) DCE-MR scans of the liver with tumor lesions, (b), quantitative parameter estimates k_{trans}, v_e and v_p for ID, ID-$data\ (\tau = 0.04)$ and OD-AIFs in-vivo data (taken from Dejene et al., 2023 [20]).

4.2.2 Uncertainty evaluation

ID in-vivo data ($\boldsymbol{\tau = 0.02}$)

Figure 4.7. and Figure 4.8. (both taken from Dejene et al., 2023 [20]) demonstrate the results for aleatoric and epistemic uncertainties for ID in-vivo data, respectively. k_{trans} and v_p showed higher aleatoric uncertainties in tumor lesions, while v_e showed smaller aleatoric uncertainties for tumor lesions. A similar behaviour for the quantitative parameter maps, for aleatoric and epistemic uncertainty patterns were obtained for the sagittal view (Figure 4.9.). Figure 4.10. (taken from Dejene et al., 2023 [20]) showed uncertainties of four additional patients with tumor lesions in the liver, estimated using the BNN. These results also demonstrated similar patterns of aleatoric and epistemic uncertainty maps for the tumor and healthy regions with the ID data of Figure 4.7.

ID in-vivo data ($\boldsymbol{\tau = 0.04}$)

ID in-vivo data with $\tau = 0.04$ have nearly the same aleatoric uncertainties as the aleatoric uncertainty of ID-data (Figure 4.7., taken from Dejene et al., 2023 [20]). The epistemic uncertainties for the quantitative parameters increased for ID in-vivo data with $\tau = 0.04$ compared to the ID in-vivo data (Figure 4.8., taken from Dejene et al., 2023 [20])

OD in-vivo data

The aleatoric and epistemic uncertainties of the quantitative parameters varied because of applying OD-AIFs to the trained BNN. Figure 4.7. (taken from Dejene et al., 2023 [20]) shows aleatoric uncertainty evaluation of OD-data. OD-AIFs showed an increase in the aleatoric uncertainty compared to the ID-data.

The results of the epistemic uncertainty evaluations are presented below (Figure 4.8., taken from Dejene et al., 2023 [20]). The epistemic uncertainties for the quantitative parameters increased more for OD-AIFs than for the ID in-vivo data.

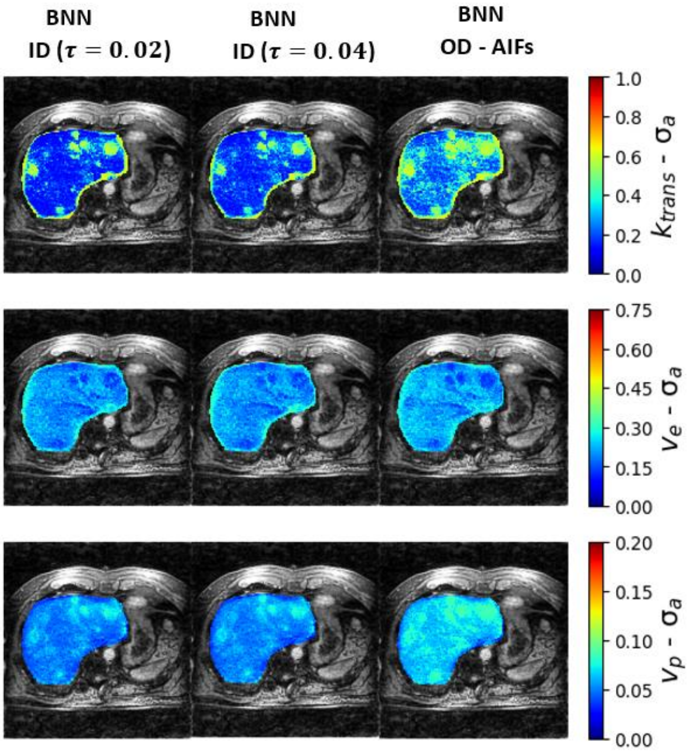


Figure 4.7. Aleatoric uncertainty estimates for ktrans, ve and vp for ID, ID-$data$ ($\tau = 0.04$) and OD-AIFs in-vivo data (taken from Dejene et al. [20], 2023).

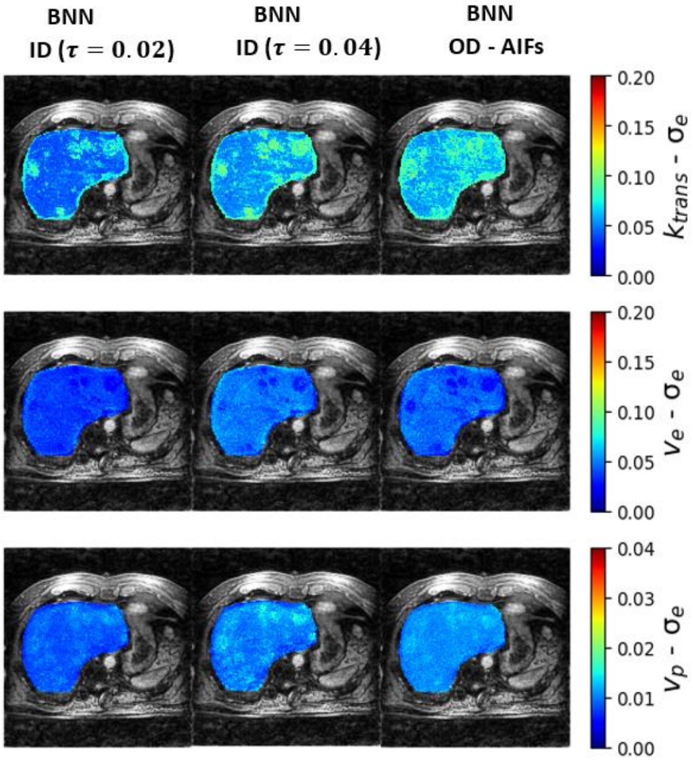


Figure 4.8. Epistemic uncertainty estimates for k_{trans}, v_e and v_p for ID, ID-$data$ ($\tau = 0.04$) and OD-AIFs in-vivo data (taken from Dejene et al., 2023 [20]).

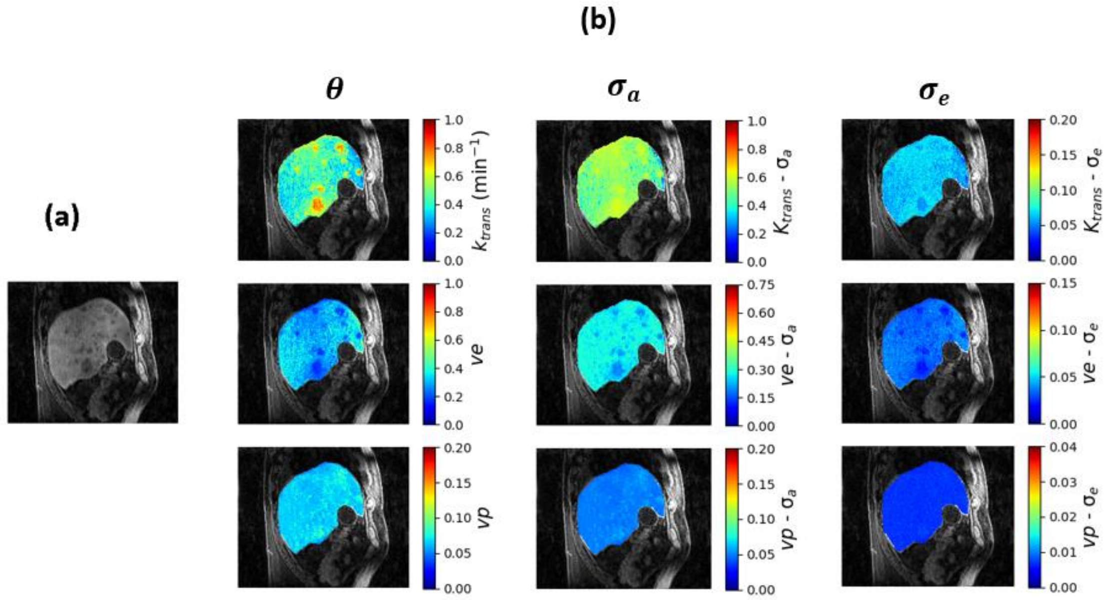


Figure 4.9. (a) DCE-MR scans of the liver with tumor lesions, (b) Quantitative parameters (θ), aleatoric uncertainties (σ_a), and epistemic uncertainties estimated for ID data.

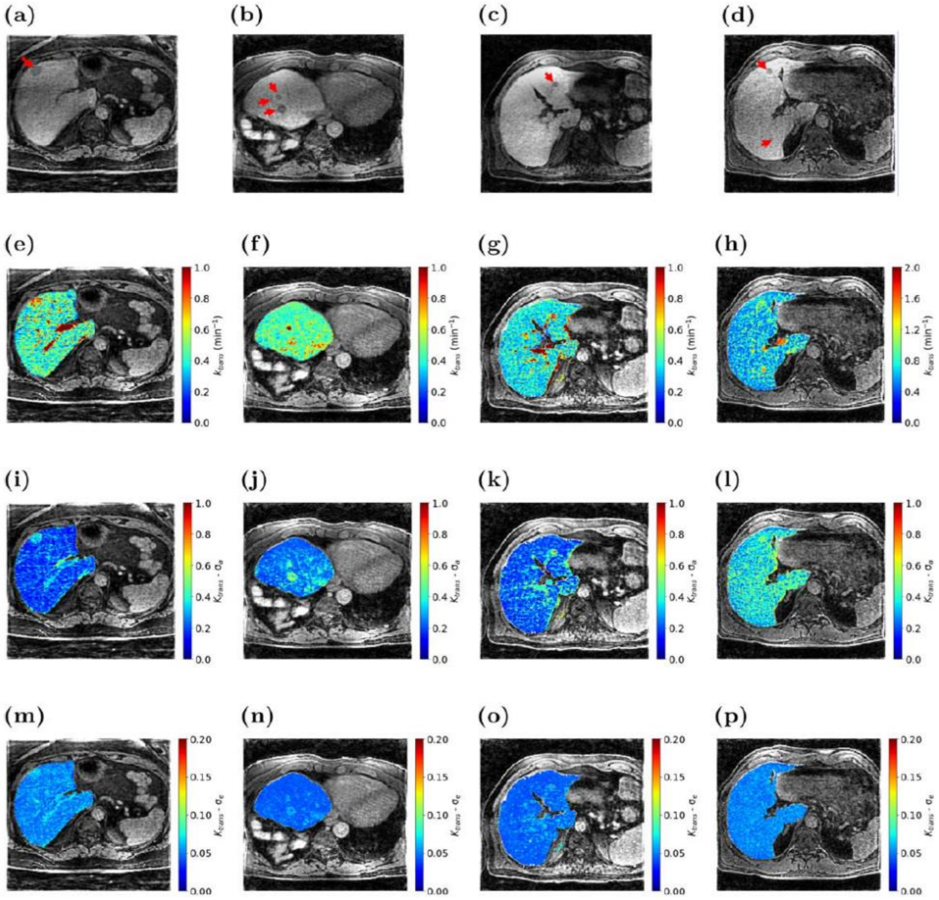


Figure 4.10. DCE images for four patients with hepatic metastasis (a-d), quantitative parameter maps (e-h), aleatoric uncertainty (i-l) and epistemic uncertainty (m-p) estimates for k_{trans} (taken from Dejene et al., 2023 [20]).

5 Discussion

In this work, we have proposed a unified Bayesian framework for uncertainty quantification of quantitative parameters from DCE-MRI scans.

5.1 Short summary of results

For the simulated data, the R^2, RMSE and AIC values demonstrated that the proposed BNN performed better than the NLLS fit for all noise levels. For the in-vivo data, parameter estimates robust to noise in the input concentration curves (Figure 4.6., from Dejene et al., 2023 [20]) were obtained for the BNN in contrary to parameter estimates by the NLLS fit, which were prone to noise (Figure 4.5., from Dejene et al., 2023 [20]). Overall, the BNN performed better in terms of accuracy of parameters, robustness to noise and computational speed.

The quantitative parameters yielded increased aleatoric uncertainties when the noise level in the time-concentration curves was increased for ID data, while the effect of noise on the epistemic uncertainty for ID data was very little. The increase in the epistemic uncertainty for ID was $0.08\% \pm 2.7\%$, $0.03\% \pm 3.4\%$ and $0.02\% \pm 5\%$ for k_{trans}, v_e and v_p on average for 0% to 15% noise levels, respectively. OD-data (i.e., OD-NLs, OD-AIFs, and OD-Δt) and less training data (ID-$\tau = 0.04$) on the other hand showed a higher epistemic uncertainty than the ID-data. OD-AIFs and OD-Δt showed higher aleatoric uncertainty and OD-NLs lower aleatoric uncertainty than the ID-data. Less training data showed very similar aleatoric uncertainties for the ID-data of both simulated and in-vivo experiments.

5.2 Interpretation of results

Parameter estimation

The quantitative parameters, k_{trans}, v_e and v_p were able to differentiate between healthy tissue and tumor lesions (Figure 4.4. and Figure 4.5., taken from Dejene et al., 2023 [20]). The contrast agent diffuses with a faster rate from the plasma to the EES compartment for the tumor lesions, which is physiologically accurate for tumor lesions characterized by massive angiogenesis and abnormal vasculature [42]. The amount of CA passing through a plasma voxel is also high for the tumours while, the percentage of contrast agent in the extravascular extracellular space (v_e) is low. Hence, tumour lesions showed high k_{trans} and v_p values and low v_e values [43]. The tumor lesions were easily detectable from the surrounding healthy tissue (Figure 4.5., taken from Dejene et al., 2023 [20]), especially with the k_{trans} and v_e maps. This showed that the obtained quantitative parameters characterized the physiology of healthy liver tissue and the pathophysiology of tumor lesions accurately.

Parameter estimation with less training data ($\tau = 0.04$) was similar with the ID-data ($\tau = 0.02$), while OD-AIF showed overestimation for k_{trans} (Figure 4.6., taken from Dejene et al., 2023 [20]) in case of in-vivo data. The experiment with less training data size had the same distribution as the training data, whereas OD-AIFs contains AIF outside the distribution of the training data. The application of OD-AIF during inference led to an overestimation of k_{trans}, which was also reflected by an increase in the respecitve aleatoric and epistemic uncertainty of the in-vivo data. Similar to studies by Huang *et al.* [16], the AIF-induced variation were larger for k_{trans} than other quantitative parameters. The quantitative parameter map patterns were unaffected by AIF-caused variations as illustrated in Figure 4.6. (taken from Dejene et al., 2023 [20]). This is also consistent with studies by Huang *et al.* [16] where the variations in the AIF did not change the quantitative parameter map patterns.

Comparison with the reference method

BNN performed better than the NLLS fit because of its higher accuracy (low RMSE, low AIC and high R^2) and robustness to noise (Figure 4.4., taken from Dejene et al., 2023 [20]), generalization ability and faster inference (0.1 min for 192 X 192 X 47 image). The generalization ability, i.e., BNN's ability to adapt to unseen in-vivo data was evident from Figure 4.4. (taken from Dejene et al., 2023 [20]). This indicated that the BNN generalized the training data, and the simulated training data also mimics in-vivo data [14].

Uncertainty

ID-data

Figure 4.1. (from Dejene et al., 2023 [20]) showed an increase in the aleatoric uncertainties with increasing noise levels. This demonstrated the sensitivity of the BNN to variations of noise levels (NL). The BNN learned the training NL (i.e., NL was ID) and was able to detect various NLs during inference. The sensitivity of the BNN to noise is in particular important for in-vivo data, in which variations in NLs can occur because of variations in patient size, amount of injected CA dose, residual under-sampling artifacts, imperfections of coil sensitivities or acquisition noise [18], [19], [17]. When in-vivo data was applied to the BNN, the noise propagated from the TCCs to the quantitative parameters was estimated by the aleatoric uncertainty.

The increase in noise levels however showed a very small effect on the epistemic uncertainty. This is expected because the test data NLs was also used for the training of the BNN, so the epistemic uncertainty did not increase. This is consistent with other studies in which the epistemic uncertainty remained unaffected for ID data [21], [36] .

OD-data

We investigated situations when less training data or OD-test data was applied to a trained BNN. The results showed the utility of calculating the epistemic uncertainty for the task of less training data and OD-data detection originating from different sources (OD-NLs, OD-AIF and OD-Δt). This was clearly shown by an increase in the epistemic uncertainty for OD data (Figure 4.3., taken from Dejene et al., 2023 [20]) in simulated and in-vivo data (Figure 4.8., taken from Dejene et al., 2023 [20]), respectively. The epistemic uncertainty increased when the training data size was reduced by 50%. This was consistent with other studies, where the epistemic uncertainty increased with OD-data or less training data size [36], [21] as a result of an increase in the variance of the posterior distribution.

For ID-$data\ (\tau = 0.04)$, although the size of the training data was reduced by half, the distribution of the application data was the same as the training data. Hence, the BNN estimated the aleatoric uncertainties for ID-$data\ (\tau = 0.04)$ correctly both in simulated (Figure 4.2., from Dejene et al., 2023 [20]) and in-vivo data (Figure 4.7., from Dejene et al., 2023 [20]). Aleatoric uncertainties for OD-AIF and OD-Δt increased, while the aleatoric uncertainty for OD-NLs decreased for simulated data. OD-AIFs for in-vivo data also showed an increased aleatoric uncertainty. This indicates that the aleatoric uncertainty could not be estimated properly for the OD-test data. This is because the test data is not in the distribution of the training data; hence, the aleatoric uncertainty estimates could not be captured by the BNN. This is also reflected by an increase in the epistemic uncertainty of the corresponding OD-data.

5.3 Embedding the results into the current state of research

The proposed framework served two main purposes: i) it provided robust quantitative parameter estimates, especially for noise-affected input concentration curves that resulted in inaccurate and imprecise results when using the standard NLLS fit method; and ii) it provided uncertainty information for the estimated quantitative parameters. Specifically, for in-vivo data, the uncertainty estimates provided additional information on the propagation of noise from the input to the quantitative parameters and the uncertainty of the trained BNN to describe the in-vivo data.

As a first novelty, the BNN provided parameter estimates more robust to noise in the DCE-MR data and more accurate as compared to the non-linear-least-squares fitting. This was validated both in numerical simulations and for in-vivo data, for which an improved performance was obtained both qualitatively and quantitatively.

Poor performance of DL-network on OD data revealed the vulnerability of DL-networks when dealing with OD data [21]. Uncertainty quantification of quantitative parameters was the second novelty of this research, which allowed the uncertainty quantification of OD cases (OD-NLs, OD-Δt and OD-AIF) for quantitative parameter estimation. Our results showed that OD-test data could be detected by the epistemic uncertainty. This could assist in clinical decision-making because the clinical decision would not only depend on the quantitative parameters but also on the uncertainties. By incorporating uncertainty, the proposed framework provided information on the mismatch between the training and application data. In the direction of estimating the epistemic uncertainty, Martin et al. [44] proposed to use the epistemic uncertainty for highlighting areas in the vessel segmentation, which requires further validation by an expert for x-ray angiograms; Nilsen et al. [45] proposed a low cost approximation of the Delta method to quantify the epistemic uncertainty for classification tasks of MNIST and CIFAR-10 datasets; Jones et al. [46] proposed to quantify the epistemic uncertainty for segmentation tasks of T1 brain images. As to our knowledge, quantification of epistemic uncertainty for DCE-MR physiological parameter estimation has not been investigated before.

The third novelty was the quantification of the epistemic uncertainty of the quantitative parameters when the training data size of the BNN was decreased. In this case, the TCCs were within the distribution of the training data, but the training data size was insufficient. This increased the epistemic uncertainty but the aleatoric uncertainty remained unaffected, indicating the ability of the proposed framework to distinguish the uncertainty arising from unseen application data and uncertainty arising from the intrinsic ambiguity of the data.

The fourth novelty of the proposed BNN was its ability to capture the propagation of noise of the input data (i.e., TCCs) to the quantitative parameters by properly estimating the aleatoric uncertainties. This is crucial to capture the propagation of noise from DCE-MR in-vivo images to the quantitative parameters.

5.4 Strengths and weaknesses of the study(s)

The proposed framework was found to succeed in providing uncertainty information in addition to the quantitative parameters. For each estimated parameter (k_{trans}, v_e and v_p), both a voxel-based aleatoric and epistemic uncertainty were estimated. This study was evaluated by using simulated data and in-vivo data from five patients with hepatic metastasis.

The strengths of the work are:

1. Generation and use of synthetic training DCE-data, which mimic the quantitative parameter and noise level ranges comparable to the distribution of the in-vivo data. Particularly, this is relevant in in-vivo cases such as for the liver, where high quality in-vivo training data are difficult to acquire because of large FOV, residual undersampling artefacts, imperfection of coil sensitivities or acquisition noise.
2. Bolus arrival time (BAT) delays between the plasma and the liver tissue was accounted for during the learning process of the DL-network, without the need for having the time delay as an additional output parameter.

3. Quantification of the epistemic uncertainty for out-of-distribution cases. One factor limiting the use of DL-networks is poor performance for unseen data. With epistemic uncertainty, we are able to quantify the reliability of quantitative parameters with respect to OD data to assist clinical decisions.
4. Separation of the predictive uncertainty into aleatoric and epistemic uncertainties.
5. Faster inference time than the conventional NLLS fitting method.

One weakness of the work is that the proposed framework does not handle input with different acquisition lengths. The training concentration curves, and the AIFs were generated using dynamic scans at 47 time points and hence, the DL-network takes an input with only 47 time points. This limits the application of the proposed network to a fixed data length. Variable input data lengths would improve the generalizability and flexibility of the proposed framework. Another weakness is that the proposed framework estimates quantitative parameters only based on temporal information and does not consider neighbourhood spatial information, which may improve the robustness of quantitative parameters to noise in the input data. Although the epistemic uncertainty was correctly estimated for OD-data, the aleatoric uncertainty for OD-data was not because the aleatoric uncertainty was learnt by the DL-network from the training data. Hence, a proper unbiased estimation of the aleatoric uncertainty for OD-data would assist to determine the contribution of noise to uncertainty apart from the data being OD. Furthermore, the study requires further testing with more patients of the same pathology to assess the benefits and limitations of the proposed framework. Further analysis of the effects of alternations in AIF, variations in noise levels, and bolus arrival time delays among different patients on quantitative parameters can assist in validating the results and assessing the future clinical applicability of the proposed method.

5.5 Implications for practice and/or future research

The motivation for this work arose from current challenges in explaining the outputs of DL-networks as mentioned by Geirhos et al. [47] and Saleem et al. [48]. We will discuss the utility of the proposed framework in clinical applications from the perspectives of performance, explanation, and computational efficiency.

The proposed BNN is robust to noise when estimating quantitative parameters as tested for various noise levels for simulated data. For in-vivo data, parameter estimates were obtained which were more robust to noise in the input concentration curves for the BNN as compared to the NLLS fit. This robustness to noise is another crucial aspect to clinical applicability to noise affected DCE-MR in-vivo patient images.

The performance of DL-networks with out-of-distribution data is one challenge for their utilization in clinical applications [47], [48]. The distribution of training data of DL-networks does not always match the distribution of the application data. The proposed framework provided a measure of the epistemic uncertainty, i.e., uncertainty of the trained network to characterize the in-vivo data. This uncertainty provided quantitative information on the performance of the BNN trained with a simulated data, when applied to in-vivo data. In future work, the application of post-hoc explainable AI (XAI) techniques [48] such as feature attribution methods, can benefit to visually assess the BNN performance when a mismatch between the training and application data occurs or, the effect of increasing noise in the time-concentration curves on the estimated quantitative parameters.

The effective clinical applicability of methods usually requires short execution times and reasonable amount of computational resources [49]. Given the DCE-MR images and AIF extracted from the patient as an input, the clinician can easily and quickly (e.g., 0.1 min for a 192 X192X47 dynamic DCE MR scan) get an estimate of both the quantitative parameters and the uncertainties by using the proposed BNN.

The training of the DL-network was performed entirely by using simulated data, meaning that it can be applied to DCE-MR images for organs different from the liver, without the need for any in-vivo training data. In the future, this framework can be adapted to other quantitative parameter estimation applications such as quantification of myocardial perfusion, blood-brain-barrier perfusion and further settings either by changing the compartmental model and/or the (patho)physiological parameter ranges applied to the simulation of the training concentration curves.

In future work, uncertainty information can be used to analyse cases where the assumption of the eTofts model is not met, e.g. when there is no diffusion of the CA from the plasma to the EES compartment. Especially in organs, such as the brain, where the blood-brain-barrier is intact to allow the passage of the CA between the compartments, epistemic uncertainty could provide the discrepancies between the training data (with assumptions of eTofts model) and the application data which do not meet the the assumptions of eTofts model.

It would be valuable to further investigate the different quantitative parameters for their ability to differentiate between different tumour types. This would assist in the treatment of different tumours and development of therapeutic measures with regard to the respective tumour type [43]. This work however requires further investigation with more patients for safe clinical applicability. There are prevailing intra-patient and inter-patient variabilities of the quantitative parameters [43] which require a broader basis of in-vivo data; and the same is true for potentially differentiating different tumour types. Future improvements of this framework may concentrate on the application of a dual input two-compartments eTofts model [38] for simulation of training data. This can assist to obtain more reliable characterization of the perfusion conditions in the liver [20]

6 Conclusions

DL-networks have been proposed for quantitative parameter estimation to overcome the limitations of the conventional NLLS fitting approaches. DL-network training however is challenged by the lack of high quality DCE-MR images, especially in organs such as the liver, because of artifacts due to (respiratory) motion, residual under-sampling artifacts, imperfection of coil sensitivities or acquisition noise. Such noise and artifacts of in-vivo images, which enter the DL-network data analysis process (e.g. during training or inference) may yield inaccurate physiological parameter estimation. In addition, it is challenging to construct an in-distribution (ID) training dataset that ensures the coverage of all possible NLs, AIFs, or TCC distributions of in-vivo data. Out-of-distribution (OD) data cases, where the training data is different from the in-vivo data, may yield inaccurate and uncertain physiological information.

In this research work, we demonstrated a novel unified Bayesian framework for estimating aleatoric and epistemic uncertainties of physiological parameters calculated from DCE-MR in-vivo imaging data of the liver for each voxel. Aleatoric uncertainty, which arises from intrinsic ambiguity of the data, was investigated with respect to noise and various levels of noise in the time-concentration curves for ID-data. Epistemic uncertainty, which arises from the discrepancy between training and application data, was investigated with respect to OD-data (OD-NLs, OD-AIFs and OD-Δt) and less training data size in comparison to the application data (ID-$data$ $(\tau = 0.04)$). In addition, the performance of NLLS fit and BNN for quantitative parameter estimation were compared, using both simulated and in-vivo data.

The quantitative parameters yielded increased aleatoric uncertainties when the noise level in the time-concentration curves was increased for ID data, while the effect of noise on the epistemic uncertainty for ID-data was very little. Less training data size and the application of OD-data increased the epistemic uncertainty. However, less training data had very little effect on the estimation of aleatoric uncertainty, while OD-data substantially influenced the accurate estimation of the aleatoric uncertainty. Thus, substantial improvement, in accuracy, robustness to noise and computational speed, was obtained by a Bayesian formulation of tracer-kinetic modelling using the proposed framework in comparison to the NLLS fit.

Reference list

[1] M. C. Hulvat, “Cancer Incidence and Trends,” *Surg. Clin. North Am.*, vol. 100, no. 3, pp. 469–481, 2020, doi: https://doi.org/10.1016/j.suc.2020.01.002.

[2] S. R. Horn, K. C. Stoltzfus, E. J. Lehrer, L. A. Dawson, L. Tchelebi, N. J. Gusani, N. K. Sharma, H. Chen, D. M. Trifiletti, and N. G. Zaorsky, “Epidemiology of liver metastases,” *Cancer Epidemiol.*, vol. 67, p. 101760, 2020, doi: 10.1016/j.canep.2020.101760.

[3] H.-Y. Zhao, Y. Gong, F.-G. Ye, H. Ling, and X. Hu, “Incidence and prognostic factors of patients with synchronous liver metastases upon initial diagnosis of breast cancer: a population-based study,” *Cancer Manag. Res.*, vol. 10, pp. 5937–5950, 2018, doi: 10.2147/CMAR.S178395.

[4] K. A. McGlynn, J. L. Petrick, and H. B. El-Serag, “Epidemiology of Hepatocellular Carcinoma,” *Hepatology*, vol. 73, no. Suppl 1, pp. 4–13, 2021, doi: 10.1002/hep.31288.

[5] R. M. Berman, A. M. Brown, S. D. Chang, S. Sankineni, M. Kadakia, B. J. Wood, P. A. Pinto, P. L. Choyke, and B. Turkbey, “DCE MRI of prostate cancer,” *Abdom. Radiol. (New York)*, vol. 41, no. 5, pp. 844–853, 2016, doi: 10.1007/s00261-015-0589-3.

[6] T. E. Yankeelov and J. C. Gore, “Dynamic Contrast Enhanced Magnetic Resonance Imaging in Oncology: Theory, Data Acquisition, Analysis, and Examples,” *Curr. Med. Imaging Rev.*, vol. 3, no. 2, pp. 91–107, 2009, doi: 10.2174/157340507780619179.

[7] K. Fang, Z. Wang, Z. Li, B. Wang, G. Han, Z. Cheng, Z. Chen, C. Lan, Y. Zhang, P. Zhao, X. Jin, Y. Liu, and R. Bai, “Convolutional neural network for accelerating the computation of the extended Tofts model in dynamic contrast-enhanced magnetic resonance imaging,” *J. Magn. Reson. Imaging*, vol. 53, no. 6, pp. 1898–1910, 2021, doi: 10.1002/jmri.27495.

[8] T. Ottens, S. Barbieri, M. R. Orton, R. Klaassen, H. W. M. van Laarhoven, H. Crezee, A. J. Nederveen, X. Zhen, and O. J. Gurney-Champion, “Deep learning DCE-MRI parameter estimation: Application in pancreatic cancer,” *Med. Image Anal.*, vol. 80, p. 102512, 2022, doi: https://doi.org/10.1016/j.media.2022.102512.

[9] D. He, X. Fan, A. Chatterjee, S. Wang, M. Medved, F. D. Pineda, A. Yousuf, T. Antic, A. Oto, and G. S. Karczmar, “A compact solution for estimation of physiological parameters from ultrafast prostate dynamic contrast enhanced MRI,” *Phys. Med. Biol.*, vol. 64, no. 15, p. 155012, 2019, doi: 10.1088/1361-6560/ab2b62.

[10] L. Chen, D. Liu, J. Zhang, B. Xie, X. Zhou, R. Grimm, X. Huang, J. Wang, and L. Feng, “Free-breathing dynamic contrast-enhanced MRI for assessment of pulmonary lesions using golden-angle radial sparse parallel imaging,” *J. Magn. Reson. Imaging*, vol. 48, no. 2, pp. 459–468, 2018, doi: 10.1002/jmri.25977.

[11] I. Polycarpou, G. Soultanidis, and C. Tsoumpas, “Synergistic motion compensation strategies for positron emission tomography when acquired simultaneously with magnetic resonance imaging,” *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 379, no. 2204, p. 20200207, 2021, doi: 10.1098/rsta.2020.0207.

[12] Y. Bliesener, J. Acharya, and K. S. Nayak, “Efficient DCE-MRI Parameter and Uncertainty Estimation Using a Neural Network,” *IEEE Trans. Med. Imaging*, vol. 39, no. 5, pp. 1712–1723, 2020, doi: 10.1109/TMI.2019.2953901.

[13] R. L. M. van Herten, A. Chiribiri, M. Breeuwer, M. Veta, and C. M. Scannell, “Physics-informed neural networks for myocardial perfusion MRI quantification,” *Med. Image Anal.*, vol. 78, p. 102399, 2022, doi: https://doi.org/10.1016/j.media.2022.102399.

[14] J. Zou, J. M. Balter, and Y. Cao, “Estimation of pharmacokinetic parameters from DCE-MRI by extracting long and short time-dependent features using an LSTM network,” *Med. Phys.*, 2020, doi: 10.1002/mp.14222.

[15] C. A. Arledge, D. M. Sankepalle, W. N. Crowe, Y. Liu, L. Wang, and D. Zhao, “Deep learning quantification of vascular pharmacokinetic parameters in mouse brain tumor models,” *Front. Biosci.*, vol. 27, no. 3, p. 99, 2022, doi: 10.31083/j.fbl2703099.

[16] W. Huang, Y. Chen, A. Fedorov, X. Li, G. H. Jajamovich, D. I. Malyarenko, M. P. Aryal, P. S. LaViolette, M. J. Oborski, F. O’Sullivan, R. G. Abramson, K. Jafari-Khouzani, A. Afzal, A. Tudorica, B. Moloney, S. N. Gupta, C. Besa, J. Kalpathy-Cramer, J. M. Mountz, C. M. Laymon, M. Muzi, K. Schmainda, Y. Cao, T. L. Chenevert, B. Taouli, T. E. Yankeelov, F. Fennessy, and X. Li, “The Impact of Arterial Input Function Determination Variations on Prostate Dynamic Contrast-Enhanced Magnetic Resonance Imaging Pharmacokinetic Modeling: A Multicenter Data Analysis Challenge,” *Tomogr. (Ann Arbor, Mich.)*, vol. 2, no. 1, pp. 56–66, 2016, doi: 10.18383/j.tom.2015.00184.

[17] H. Kim, “Modification of population based arterial input function to incorporate individual variation,” *Magn. Reson. Imaging*, vol. 45, pp. 66–71, 2018, doi: 10.1016/j.mri.2017.09.010.

[18] H. Jiao, X. Jiang, Z. Pang, X. Lin, Y. Huang, and L. Li, “Deep Convolutional Neural Networks-Based Automatic Breast Segmentation and Mass Detection in DCE-MRI,” *Comput. Math. Methods Med.*, vol. 2020, p. 2413706, 2020, doi: 10.1155/2020/2413706.

[19] M. Ippoliti, M. Lukas, W. Brenner, I. Schatka, C. Furth, T. Schaeffter, M. R. Makowski, and C. Kolbitsch, “Respiratory motion correction for enhanced quantification of hepatic lesions in simultaneous PET and DCE-MR imaging,” *Phys. Med. Biol.*, vol. 66, no. 9, p. 95012, 2021, doi: 10.1088/1361-6560/abf51e.

[20] E. M. Dejene, W. Brenner, M. R. Makowski, and C. Kolbitsch, “Unified Bayesian network for uncertainty quantification of physiological parameters in dynamic contrast enhanced (DCE) MRI of the liver,” *Phys. Med. Biol.*, vol. 68, no. 21, p. 215018, 2023, doi: 10.1088/1361-6560/ad0284.

[21] M. Valdenegro-Toro and D. S. Mori, “A Deeper Look into Aleatoric and Epistemic Uncertainty Disentanglement,” in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2022, pp. 1508–1516. doi: 10.1109/CVPRW56347.2022.00157.

[22] M. Medved, G. Karczmar, C. Yang, J. Dignam, T. F. Gajewski, H. Kindler, E. Vokes, P. MacEneany, M. T. Mitchell, and W. M. Stadler, “Semiquantitative analysis of dynamic contrast enhanced MRI in cancer patients: Variability and changes in tumor tissue over time,” *J. Magn. Reson. Imaging*, vol. 20, no. 1, pp. 122–128, 2004, doi: https://doi.org/10.1002/jmri.20061.

[23] P. S. Tofts, “Modeling tracer kinetics in dynamic Gd-DTPA MR imaging,” *J. Magn. Reson. Imaging*, vol. 7, no. 1, pp. 91–101, 1997, doi: 10.1002/jmri.1880070113.

[24] A. L. Bendinger, C. Debus, C. Glowa, C. P. Karger, J. Peter, and M. Storath, “Bolus arrival time estimation in dynamic contrast-enhanced magnetic resonance imaging of small animals based on spline models,” *Phys. Med. Biol.*, vol. 64, no. 4, p. 45003, 2019, doi: 10.1088/1361-6560/aafce7.

[25] A. Singh, R. K. S. Rathore, M. Haris, S. K. Verma, N. Husain, and R. K. Gupta, “Improved bolus arrival time and arterial input function estimation for tracer kinetic analysis in DCE-MRI,” *J. Magn. Reson. Imaging*, vol. 29, no. 1, pp. 166–176, 2009, doi: 10.1002/jmri.21624.

[26] H. Abdel-Jaber, D. Devassy, A. Al Salam, L. Hidaytallah, and M. EL-Amir, “A Review of Deep Learning Algorithms and Their Applications in Healthcare,” *Algorithms*, vol. 15, no. 2, 2022, doi: 10.3390/a15020071.

[27] S. R. Dubey, S. K. Singh, and B. Chaudhuri, “Activation Functions in Deep Learning: A comprehensive Survey and Benchmark,” *Neurocomputing*, vol. 503, 2022, doi: 10.1016/j.neucom.2022.06.111.

[28] D. M. Olsson and L. S. Nelson, “The Nelder-Mead Simplex Procedure for Function Minimization,” *Technometrics*, vol. 17, no. 1, pp. 45–51, 1975, doi: 10.1080/00401706.1975.10489269.

[29] R. Saravanan and P. Sujatha, “A State of Art Techniques on Machine Learning Algorithms: A Perspective of Supervised Learning Approaches in Data Classification,” in *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2018, pp. 945–949. doi: 10.1109/ICCONS.2018.8663155.

[30] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, V. Makarenkov, and S. Nahavandi, “A review of uncertainty quantification in deep learning: Techniques, applications and challenges,” *Inf. Fusion*, vol. 76, pp. 243–297, 2021, doi: https://doi.org/10.1016/j.inffus.2021.05.008.

[31] F. Glang, A. Deshmane, S. Prokudin, F. Martin, K. Herz, T. Lindig, B. Bender, K. Scheffler, and M. Zaiss, “DeepCEST 3T: Robust MRI parameter determination and uncertainty quantification with neural networks-application to CEST imaging of the human brain at 3T,” *Magn. Reson. Med.*, vol. 84, no. 1, pp. 450–466, 2020, doi: 10.1002/mrm.28117.

[32] C. F. Westbury, “Bayes’ rule for clinicians: an introduction,” *Front. Psychol.*, vol. 1, p. 192, 2010, doi: 10.3389/fpsyg.2010.00192.

[33] D. Tran, M. Dusenberry, M. van der Wilk, and D. Hafner, “Bayesian Layers: A Module for Neural Network Uncertainty,” in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d\textquotesingle Alché-Buc, E. Fox, and R. Garnett, Eds., Curran Associates, Inc., 2019. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2019/file/154ff8944e6eac05d0675c95b5b8889d-Paper.pdf

[34] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009.

[35] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: A System for Large-Scale Machine Learning,” in *Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation*, in OSDI’16. USA: USENIX Association, 2016, pp. 265–283.

[36] E. Hüllermeier and W. Waegeman, “Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods,” *Mach. Learn.*, vol. 110, no. 3, pp. 457–506, 2021, doi: 10.1007/s10994-021-05946-3.

[37] S. Miyazaki, K. Murase, T. Yoshikawa, S. Morimoto, Y. Ohno, and K. Sugimura, “A quantitative method for estimating hepatic blood flow using a dual-input single-compartment model,” *Br. J. Radiol.*, vol. 81, no. 970, pp. 790–800, 2008, doi: 10.1259/bjr/52166324.

[38] M. D. Chouhan, A. Bainbridge, D. Atkinson, S. Punwani, R. P. Mookerjee, M. F. Lythgoe, and S. A. Taylor, “Estimation of contrast agent bolus arrival delays for improved

reproducibility of liver DCE MRI," *Phys. Med. Biol.*, vol. 61, no. 19, pp. 6905–6918, 2016, doi: 10.1088/0031-9155/61/19/6905.

[39] B. Callewaert, E. A. V Jones, U. Himmelreich, and W. Gsell, "Non-Invasive Evaluation of Cerebral Microvasculature Using Pre-Clinical MRI: Principles, Advantages and Limitations," *Diagnostics*, vol. 11, no. 6, 2021, doi: 10.3390/diagnostics11060926.

[40] M. Ippoliti, M. Lukas, W. Brenner, T. Schaeffter, M. R. Makowski, and C. Kolbitsch, "3D nonrigid motion correction for quantitative assessment of hepatic lesions in DCE-MRI," *Magn. Reson. Med.*, vol. 82, no. 5, pp. 1753–1766, Nov. 2019, doi: https://doi.org/10.1002/mrm.27867.

[41] H. Bozdogan, "Model Selection and Akaike's Information Criterion (AIC): The General Theory and Its Analytical Extensions," *Psychometrika*, vol. 52, pp. 345–370, 1987, doi: 10.1007/BF02294361.

[42] F. U. Shen, J. Lu, L. Chen, Z. Wang, and Y. Chen, "Diagnostic value of dynamic contrast-enhanced magnetic resonance imaging in rectal cancer and its correlation with tumor differentiation," *Mol. Clin. Oncol.*, vol. 4, no. 4, pp. 500–506, 2016, doi: 10.3892/mco.2016.762.

[43] R. A. Little, H. Barjat, J. I. Hare, M. Jenner, Y. Watson, S. Cheung, K. Holliday, W. Zhang, J. P. B. O'Connor, S. T. Barry, S. Puri, G. J. M. Parker, and J. C. Waterton, "Evaluation of dynamic contrast-enhanced MRI biomarkers for stratified cancer medicine: How do permeability and perfusion vary between human tumours?," *Magn. Reson. Imaging*, vol. 46, pp. 98–105, 2018, doi: https://doi.org/10.1016/j.mri.2017.11.008.

[44] R. Martin, J. Miro, and L. Duong, "Epistemic Uncertainty Modeling for Vessel Segmentation," *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.*, vol. 2019, pp. 5923–5927, 2019, doi: 10.1109/EMBC.2019.8857785.

[45] G. K. Nilsen, A. Z. Munthe-Kaas, H. J. Skaug, and M. Brun, "Epistemic uncertainty quantification in deep learning classification by the Delta method," *Neural Netw.*, vol. 145, pp. 164–176, 2022, doi: 10.1016/j.neunet.2021.10.014.

[46] C. K. Jones, G. Wang, V. Yedavalli, and H. Sair, "Direct quantification of epistemic and aleatoric uncertainty in 3D U-net segmentation," *J. Med. imaging (Bellingham, Wash.)*, vol. 9, no. 3, p. 34002, 2022, doi: 10.1117/1.JMI.9.3.034002.

[47] R. Geirhos, J.-H. Jacobsen, C. Michaelis, R. Zemel, W. Brendel, M. Bethge, and F. A. Wichmann, "Shortcut learning in deep neural networks," *Nat. Mach. Intell.*, vol. 2, no. 11, pp. 665–673, 2020, doi: 10.1038/s42256-020-00257-z.

[48] R. Saleem, B. Yuan, F. Kurugollu, A. Anjum, and L. Liu, "Explaining deep neural networks: A survey on the global interpretation methods," *Neurocomputing*, vol. 513, pp. 165–180, 2022, doi: https://doi.org/10.1016/j.neucom.2022.09.129.

[49] M. Salvador, F. Regazzoni, L. Dede', and A. Quarteroni, "Fast and robust parameter estimation with uncertainty quantification for the cardiac function," *Comput. Methods Programs Biomed.*, vol. 231, p. 107402, 2023, doi: 10.1016/j.cmpb.2023.107402.

Statutory Declaration

"I, Edengenet Mashilla Dejene, by personally signing this document in lieu of an oath, hereby affirm that I prepared the submitted dissertation on the topic Bayesian Framework for Aleatoric and Epistemic Uncertainty Quantification of Quantitative Parameters: Application to DCE-MRI of the Liver / Bayes'scher Ansatz zur Quantifizierung der aleatorischen und epistemischen Unsicherheit von quantitativen Parametern: Anwendung bei der DCE-MRT der Leber, independently and without the support of third parties, and that I used no other sources and aids than those stated.
All parts which are based on the publications or presentations of other authors, either in letter or in spirit, are specified as such in accordance with the citing guidelines. The sections on methodology (in particular regarding practical work, laboratory regulations, statistical processing) and results (in particular regarding figures, charts and tables) are exclusively my responsibility.

Furthermore, I declare that I have correctly marked all of the data, the analyses, and the conclusions generated from data obtained in collaboration with other persons, and that I have correctly marked my own contribution and the contributions of other persons (cf. declaration of contribution). I have correctly marked all texts or parts of texts that were generated in collaboration with other persons.

My contributions to any publications to this dissertation correspond to those stated in the below joint declaration made together with the supervisor. All publications created within the scope of the dissertation comply with the guidelines of the ICMJE (International Committee of Medical Journal Editors; http://www.icmje.org) on authorship. In addition, I declare that I shall comply with the regulations of Charité – Universitätsmedizin Berlin on ensuring good scientific practice.

I declare that I have not yet submitted this dissertation in identical or similar form to another Faculty.

The significance of this statutory declaration and the consequences of a false statutory declaration under criminal law (Sections 156, 161 of the German Criminal Code) are known to me."

Date Signature

Declaration of your own contribution to the publications

Edengenet Mashilla Dejene contributed the following to the below listed publication:

Publication:

Dejene EM, Brenner W, Makowski MR, Kolbitsch C. Unified Bayesian network for uncertainty quantification of physiological parameters in dynamic contrast enhanced (DCE) MRI of the liver. *Phys Med Biol*. 2023;68(21)

Contribution:

- **Concept** (Edengenet Dejene, Christoph Kolbitsch): The general idea was initially proposed by Christoph Kolbistch. Edengenet Dejene formulated the idea and methodology and structured the publication with the support of Christoph Kolbistch.
- **Data acquisition and patient recruitment process** (Winfried Brenner, Markus Makowski, Christoph Kolbitsch): Winfried Brenner, Markus Makowski and Christoph Kolbitsch provided the DCE-MR in-vivo images.
- **Data simulation** (Edengenet Dejene): Edengenet Dejene simulated DCE-MR data for training and validation of the proposed framework.
- **Method implementation** (Edengenet Dejene): Edengenet Dejene implemented the code for the DL-network (BNN), implemented the standard non-linear-least-squares (NLLS) fitting in python.
- **Image processing and visualization** (Edengenet Dejene): Edengenet Dejene implemented the code for image analysis (both for simulation and in-vivo data), visualization such as, generation of tables, figures used in the dissertation and publication using python. All figures (Figures 1 to Figure 11) and all tables (Table 1 to Table 3) in the publication were created by Edengenet Dejene.
- **Development of the hypothesis** (Edengenet Dejene, Christoph Kolbitsch): Hypothesis was developed by Edengenet Dejene with the support of Christoph Kolbitsch. Edengenet Dejene investigated the hypothesis with the analysis of numerical simulations and in-vivo images.

- **Data Interpretation and Analysis** (Edengenet Dejene, Christoph Kolbitsch): The development and implementation of the BNN was performed by Edengenet Dejene. Christoph Kolbitsch was involved in the design of the experimental setup, discussions about interpretations and evaluations of the results. The implementation of the BNN resulted in estimates of physiological parameters and their corresponding uncertainties shown in Figure 4.6., Figure 4.7. and Figure 4.8. Edengenet Dejene, in close collaboration with Christoph Kolbitsch was able to demonstrate a link between aleatoric uncertainty and noise inherent to the data (Figure 4.1.), epistemic uncertainty and less training data size (Figure 4.2.) and out-of-distribution (OD) data (Figure 4.3.).
- **Clinical evaluation** (Edengenet Dejene, Christoph Kolbitsch, Winfried Brenner): Edengenet Dejene, Christoph Kolbitsch and Winfried Brenner were involved in clinical evaluation of the results.
- **Manuscript writing** (Edengenet Dejene): The first draft was written by Edengenet Dejene and revised by Christoph Kolbitsch. All co-authors have revised the final manuscript.

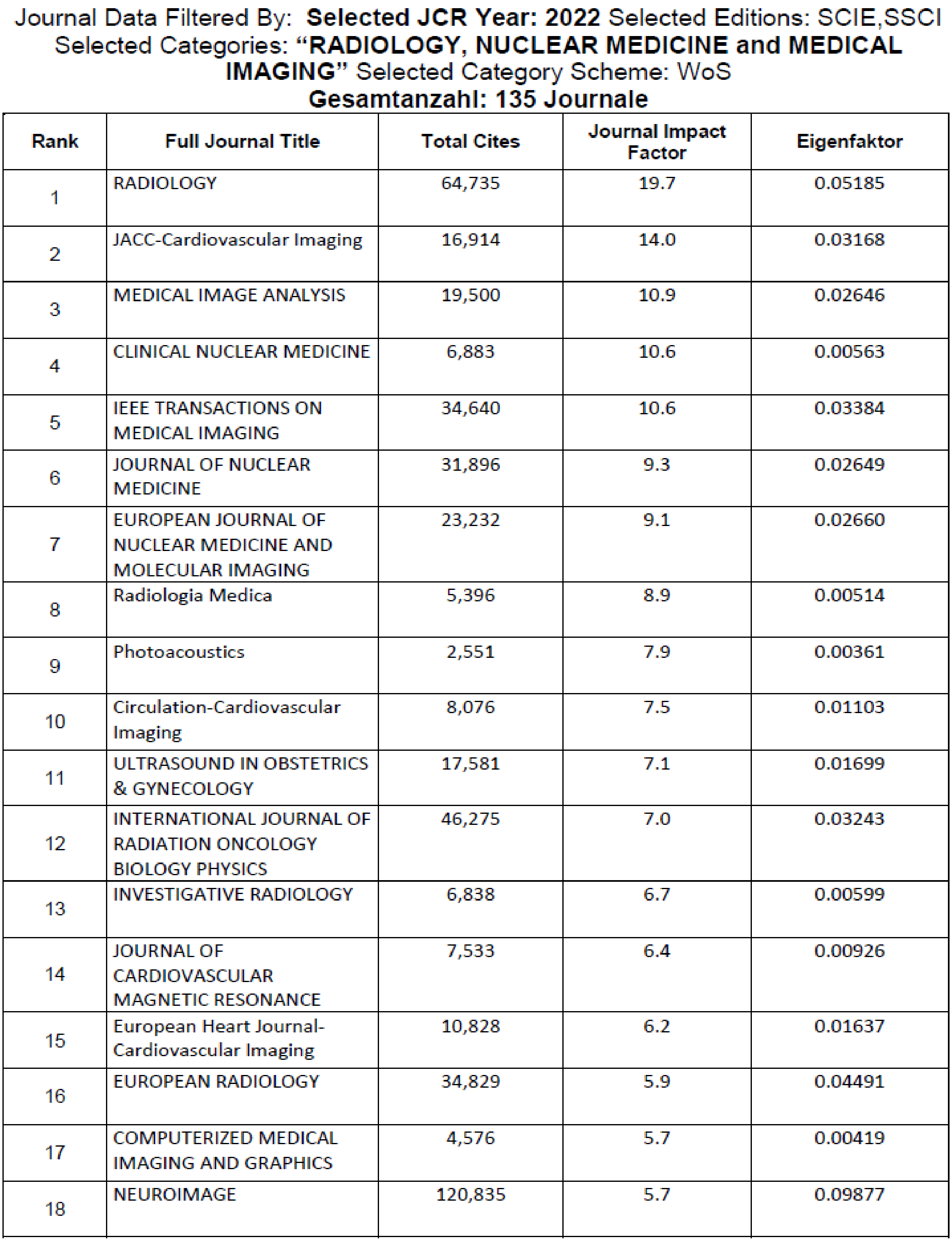
____

Signature, date and stamp of first supervising university professor / lecturer

____

Signature of doctoral candidate

Excerpt from Journal Summary List



Journal Data Filtered By: **Selected JCR Year: 2022** Selected Editions: SCIE,SSCI Selected Categories: **“RADIOLOGY, NUCLEAR MEDICINE and MEDICAL IMAGING”** Selected Category Scheme: WoS

Gesamtanzahl: 135 Journale

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfaktor
1	RADIOLOGY	64,735	19.7	0.05185
2	JACC-Cardiovascular Imaging	16,914	14.0	0.03168
3	MEDICAL IMAGE ANALYSIS	19,500	10.9	0.02646
4	CLINICAL NUCLEAR MEDICINE	6,883	10.6	0.00563
5	IEEE TRANSACTIONS ON MEDICAL IMAGING	34,640	10.6	0.03384
6	JOURNAL OF NUCLEAR MEDICINE	31,896	9.3	0.02649
7	EUROPEAN JOURNAL OF NUCLEAR MEDICINE AND MOLECULAR IMAGING	23,232	9.1	0.02660
8	Radiologia Medica	5,396	8.9	0.00514
9	Photoacoustics	2,551	7.9	0.00361
10	Circulation-Cardiovascular Imaging	8,076	7.5	0.01103
11	ULTRASOUND IN OBSTETRICS & GYNECOLOGY	17,581	7.1	0.01699
12	INTERNATIONAL JOURNAL OF RADIATION ONCOLOGY BIOLOGY PHYSICS	46,275	7.0	0.03243
13	INVESTIGATIVE RADIOLOGY	6,838	6.7	0.00599
14	JOURNAL OF CARDIOVASCULAR MAGNETIC RESONANCE	7,533	6.4	0.00926
15	European Heart Journal-Cardiovascular Imaging	10,828	6.2	0.01637
16	EUROPEAN RADIOLOGY	34,829	5.9	0.04491
17	COMPUTERIZED MEDICAL IMAGING AND GRAPHICS	4,576	5.7	0.00419
18	NEUROIMAGE	120,835	5.7	0.09877

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfaktor
19	RADIOTHERAPY AND ONCOLOGY	23,221	5.7	0.02342
20	Diagnostic and Interventional Imaging	3,017	5.5	0.00310
21	RADIOGRAPHICS	17,226	5.5	0.01063
22	Journal of Cardiovascular Computed Tomography	2,968	5.4	0.00428
23	AMERICAN JOURNAL OF ROENTGENOLOGY	35,480	5.0	0.01980
24	CANCER IMAGING	2,779	4.9	0.00307
25	SEMINARS IN NUCLEAR MEDICINE	2,877	4.9	0.00198
26	ACADEMIC RADIOLOGY	8,982	4.8	0.00989
27	HUMAN BRAIN MAPPING	28,764	4.8	0.03181
28	KOREAN JOURNAL OF RADIOLOGY	5,377	4.8	0.00639
29	Insights into Imaging	5,008	4.7	0.00700
30	Journal of the American College of Radiology	7,214	4.5	0.01336
31	JOURNAL OF DIGITAL IMAGING	5,216	4.4	0.00545
32	JOURNAL OF MAGNETIC RESONANCE IMAGING	20,316	4.4	0.01883
33	Physical and Engineering Sciences in Medicine	1,025	4.4	0.00131
34	ULTRASONICS	9,972	4.2	0.00619
35	EJNMMI Physics	1,273	4.0	0.00205
36	MEDICAL PHYSICS	31,491	3.8	0.02232
37	Radiation Oncology	8,845	3.6	0.00929
38	AMERICAN JOURNAL OF NEURORADIOLOGY	25,993	3.5	0.01811
39	JOURNAL OF BIOMEDICAL OPTICS	14,729	3.5	0.00790

Rank	Full Journal Title	Total Cites	Journal Impact Factor	Eigenfaktor
40	JOURNAL OF NEURORADIOLOGY	1,505	3.5	0.00183
41	PHYSICS IN MEDICINE AND BIOLOGY	29,272	3.5	0.01938
42	SEMINARS IN RADIATION ONCOLOGY	2,801	3.5	0.00226
43	Biomedical Optics Express	14,699	3.4	0.01758
44	CANCER BIOTHERAPY AND RADIOPHARMACEUTICALS	2,857	3.4	0.00221
45	Physica Medica-European Journal of Medical Physics	5,488	3.4	0.00739
46	RADIATION RESEARCH	9,234	3.4	0.00370
47	ULTRASCHALL IN DER MEDIZIN	2,686	3.4	0.00262
48	DENTOMAXILLOFACIAL RADIOLOGY	4,025	3.3	0.00231
49	EUROPEAN JOURNAL OF RADIOLOGY	17,302	3.3	0.01536
50	JOURNAL OF THORACIC IMAGING	2,018	3.3	0.00199
51	MAGNETIC RESONANCE IN MEDICINE	32,434	3.3	0.02128
52	Practical Radiation Oncology	2,793	3.3	0.00482
53	EJNMMI Research	2,875	3.2	0.00429
54	CANADIAN ASSOCIATION OF RADIOLOGISTS JOURNAL-JOURNAL DE L ASSOCIATION CANADIENNE DES RADIOLOGISTES	1,513	3.1	0.00185
55	Clinical and Translational Radiation Oncology	1,540	3.1	0.00325
56	INTERNATIONAL JOURNAL OF HYPERTHERMIA	6,456	3.1	0.00520
57	MOLECULAR IMAGING AND BIOLOGY	3,530	3.1	0.00407
58	NUCLEAR MEDICINE AND BIOLOGY	3,714	3.1	0.00158
59	STRAHLENTHERAPIE UND ONKOLOGIE	3,580	3.1	0.00298

Printing copy(s) of the publication(s)

Dejene EM, Brenner W, Makowski MR, Kolbitsch C. Unified Bayesian network for uncertainty quantification of physiological parameters in dynamic contrast enhanced (DCE) MRI of the liver. *Phys Med Biol*. 2023;68(21)

https://iopscience.iop.org/article/10.1088/1361-6560/ad0284

IOP Publishing Phys. Med. Biol. 68 (2023) 215018 https://doi.org/10.1088/1361-6560/ad0284

Physics in Medicine & Biology

RECEIVED
6 February 2023

REVISED
5 October 2023

ACCEPTED FOR PUBLICATION
11 October 2023

PUBLISHED
1 November 2023

PAPER

Unified Bayesian network for uncertainty quantification of physiological parameters in dynamic contrast enhanced (DCE) MRI of the liver

Edengenet M Dejene[1,2], **Winfried Brenner**[3], **Marcus R Makowski**[2,4] and **Christoph Kolbitsch**[1]

[1] Physikalisch-Technische Bundesanstalt (PTB), Braunschweig and Berlin, Germany
[2] Department of Radiology, Charité Universitätsmedizin Berlin, Berlin, Germany
[3] Department of Nuclear Medicine, Charité Universitätsmedizin Berlin, Berlin, Germany
[4] Department of Diagnostic and Interventional Radiology, Faculty of Medicine, Technical University of Munich, Munich, Germany

E-mail: edengenet.dejene@ptb.de

Keywords: quantitative imaging, DCE MRI, liver perfusion, tracer kinetic modeling, parameter estimation, uncertainty quantification, Bayesian neural networks

Abstract

Objective. Physiological parameter estimation is affected by intrinsic ambiguity in the data such as noise and model inaccuracies. The aim of this work is to provide a deep learning framework for accurate parameter and uncertainty estimates for DCE-MRI in the liver. *Approach.* Concentration time curves are simulated to train a Bayesian neural network (BNN). Training of the BNN involves minimization of a loss function that jointly minimizes the aleatoric and epistemic uncertainties. Uncertainty estimation is evaluated for different noise levels and for different out of distribution (OD) cases, i.e. where the data during inference differs strongly to the data during training. The accuracy of parameter estimates are compared to a nonlinear least squares (NLLS) fitting in numerical simulations and *in vivo* data of a patient suffering from hepatic tumor lesions. *Main results.* BNN achieved lower root-mean-squared-errors (RMSE) than the NLLS for the simulated data. RMSE of BNN was on overage of all noise levels lower by 33% ± 1.9% for k_{trans}, 22% ± 6% for v_e and 89% ± 5% for v_p than the NLLS. The aleatoric uncertainties of the parameters increased with increasing noise level, whereas the epistemic uncertainty increased when a BNN was evaluated with OD data. For the *in vivo* data, more robust parameter estimations were obtained by the BNN than the NLLS fit. In addition, the differences between estimated parameters for healthy and tumor regions-of-interest were significant ($p < 0.0001$). *Significance.* The proposed framework allowed for accurate parameter estimates for quantitative DCE-MRI. In addition, the BNN provided uncertainty estimates which highlighted cases of high noise and in which the training data did not match the data during inference. This is important for clinical application because it would indicate cases in which the trained model is inadequate and additional training with an adapted training data set is required.

1. Introduction

Dynamic contrast-enhanced MR (DCE-MR) imaging has been used as a non-invasive *in vivo* imaging modality for diagnosis and treatment monitoring of cancer by injection of a gadolinium (Gd) based contrast agent (CA) (Choyke *et al* 2003, Jackson *et al* 2007, Gurney-Champion *et al* 2020, Dündar *et al* 2022). A series of dynamic T1-weighted images are then acquired to capture the temporal changes in the CA concentration in the tissue.

Quantitative analysis of temporal changes (concentration time curves, CTCs) with tracer kinetic modeling enables the estimation of quantitative physiological parameters, such as vascular permeability and tissue perfusion (Heye *et al* 2016, Fang *et al* 2021, Ottens *et al* 2022). These parameters have been shown to improve diagnosis and treatment planning of various diseases (Berks *et al* 2021, Wang *et al* 2022). Nonlinear least squares (NLLS) fit (Ahearn *et al* 2005) is conventionally used for the fitting of tracer kinetic models to the measured

---

CTCs to estimate the physiological parameters. NLLS fitting suffers from high computational cost (Fang *et al* 2021), is strongly dependent on parameter initialization (Bliesener *et al* 2020), yields parameter estimates with large variance and bias (Ottens *et al* 2022), and the proposed solution can converge to local minimas leading to inaccurate results (Rudolf *et al* 2022). In addition, intrinsic noise in the data (Bliesener *et al* 2020) and errors when measuring the arterial input function (AIF) (Klepaczko *et al* 2020) can introduce inaccuracies and degrade precision of parameter estimates.

Deep learning (DL) has been shown to improve DCE parameter estimation compared to the NLLS method due to its generalization ability and faster inference (Ottens *et al* 2022). Several DL-based parameter estimation methods have been proposed (Ulas *et al* 2019, Klepaczko *et al* 2020, Zou *et al* 2020, Fang *et al* 2021, Ottens *et al* 2022, Rudolf *et al* 2022). These approaches only estimate point estimates of tracer kinetic parameters and do not provide any information about how reliable the parameter estimates are.

Bliesener *et al* proposed a DL-based approach to estimate parameters and aleatoric uncertainties, i.e. uncertainties in the estimated parameters due to noise in the data (Bliesener *et al* 2020). They did not assess epistemic uncertainties, i.e. uncertainties which stem from a mismatch between data used during training and inference. Epistemic uncertainties are especially interesting for 3D DCE-MRI in the liver, because high quality *in vivo* data for training is not available due to the large field-of-view and scan time limitations due to respiratory motion. Therefore, only simulated data can be used for training. If the data used for training is very different to the *in vivo* data where the trained model is applied to, wrong parameter estimates can occur. Therefore, it is important to have an indicator of the reliability of the outcome of the network to detect such out-of-distribution (OD) cases. This indicator can then be used to identify patients where the trained model is not suitable and needs to be adapted. In addition, estimating the AIF can be challenging and hence differences between estimated and the true AIF also need to be detected to ensure reliable parameter estimates.

In this paper, we present a Bayesian neural network (BNN) framework for accurate estimation of perfusion parameters for 3D DCE-MRI of the liver and quantify the aleatoric and epistemic uncertainties associated with these parameters on a pixel-by-pixel basis for hepatic tumor characterization. We investigate the uncertainty estimation in two cases: (i) capturing of aleatoric uncertainties related to noise and variations of noise levels in the data (ii) estimation of epistemic uncertainty resulting from noise levels, CTCs, inaccurate estimation of the AIF or uncorrected delays in the bolus arrival, which were OD. The framework is evaluated in numerical simulations and applied to patients suffering from hepatic tumor lesions.

2. Methods

In this section, we describe the proposed parameter estimation and uncertainty quantification framework. The core of our approach is a BNN that employs an input-dependent heteroscedastic noise model for parameter estimation and capturing of aleatoric uncertainty. A variational approximation technique is used for modeling epistemic uncertainty. The parameters of the BNN are learned based on simulated data. The data is simulated using an extended Tofts (*e*Tofts) model (Tofts *et al* 1999) and AIFs obtained from *in vivo* patients DCE scan. The structure of the framework is shown in figure 1.

2.1. BNN architecture

The BNN takes the one-dimensional CA concentration in the plasma, $C_p(t)$ and the CA concentration in the tissue, $C(t)$ simulated with the corresponding $C_p(t)$, as inputs. A multiscale temporal filter bank (Bliesener *et al* 2020) was applied to each of the inputs $C_p(t)$ and $C(t)$ separately to intrinsically learn the relation between $C_p(t)$ and $C(t)$. Each temporal filter bank consists of three 1D convolutional layers with different filter length (i.e. 10, 5 and 3) and stride length (10, 1 and 1) to extract low-level, medium-level and high-level temporal features, respectively.

The pre-connected temporal filter banks to the $C_p(t)$ and $C(t)$ with high, medium and small filter sizes are intended to extract features with high, medium and low temporal resolution from each of the inputs, respectively. To observe the fast dynamics of the flow of the CA at the early phase, a filter with the smallest size and stride combination was used. This provides a high-temporal resolution information to capture the flow (Cristina 2015, Bliesener *et al* 2020). With medium temporal resolution, i.e. medium filter size, longer time scales could be observed. This enables the estimation of ve as the backflux of the contrast agents happens on longer time scales (Cuenod and Balvay 2013, Hansen *et al* 2019, Bliesener *et al* 2020). On the other hand, the filters with the highest filter and stride size provide a low temporal resolution information. In low temporal resolutions, contrast leakages could not be observed as they require longer time periods. The total tissue concentration, $C(t)$, which is calculated as $C(t) = v_pC_p(t) + v_eC_e(t)$ becomes $C(t) = v_pC_p(t)$ when there are no contrast leakages (Sourbron and Buckley 2013, Bliesener *et al* 2020). Hence, the ratio of $C(t)$ and $C_p(t)$ for the low temporal resolution filters allow computation of v_p when there are no contrast leakages. The outputs of the

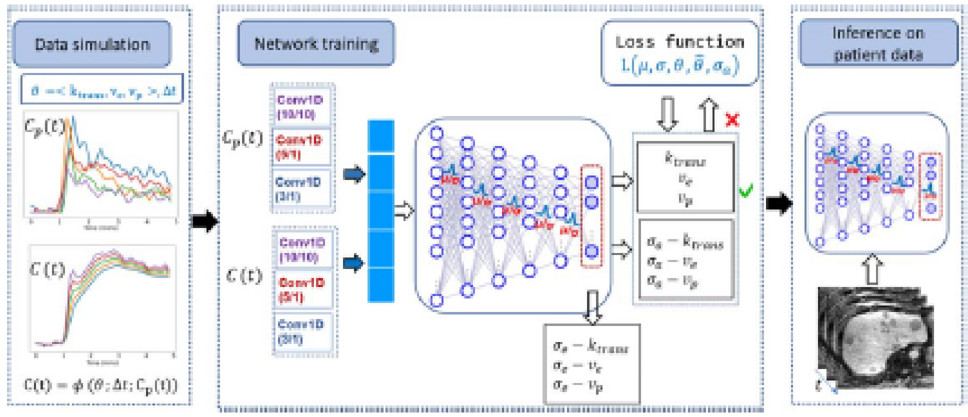


Figure 1. The proposed deep learning framework for parameter estimation and uncertainty quantification. The extended Tofts (eTofts) model, ϕ, is applied to the physiological parameters of the eTofts model, $\theta = \{k_{\mathrm{trans}}, v_e \text{ and } v_p\}$, time delay, Δt and time-dependent CA concentration in the plasma, $C_p(t)$ to simulate the time-dependent CA concentration in the tissue, $C(t)$. The BNN takes $C(t)$ and $C_p(t)$ as input to estimate parameters of the eTofts model $\hat{\theta}$, aleatoric uncertainties, $\sigma_a = \{\sigma_a - k_{\mathrm{trans}}, \sigma_a - v_e \text{ and } \sigma_a - v_p\}$ and epistemic uncertainties, $\sigma_e = \{\sigma_e - k_{\mathrm{trans}}, \sigma_e - v_e \text{ and } \sigma_e - v_p\}$ by minimizing the loss function L. μ and σ, which are the mean and standard deviation of a Gaussian distribution, are updated iteratively together with $\hat{\theta}$ and σ_a for the minimization of L.

extracted temporal features from each of the temporal filter banks are concatenated to fuse input information of $C_p(t)$ and $C(t)$ as shown in figure 1.

The BNN consists of six fully connected layers as shown in figure 1 which have 600, 400, 300, 200 and 100 and 6 neurons in each of the layers. The network architecture in Bliesener *et al* (2020) was adopted and customized into a BNN with two additional layers and output neurons specific to the application of the proposed approach. The network depth was increased to six layers to proportionally distribute neurons in the third, fourth and fifth layers. Leaky rectified linear units (RELU) activation functions were used for the first five layers to introduce nonlinearity into the mapping and the last layer has sigmoid activation function to yield parameters in physiological ranges of 0 and 1. To ensure that the estimated k_{trans} values lay in physiological ranges, the sigmoid activation function for k_{trans} estimation was customized pre-activation to yield parameter values in ranges of 0 and 2.

The outputs of the BNN are the posterior distribution of three parameters of the eTofts model $\theta = \{k_{\mathrm{trans}}, v_e$ and $v_p\}$ and their aleatoric uncertainty $\sigma_a = \{\sigma_a - k_{\mathrm{trans}}, \sigma_a - v_e$ and $\sigma_a - v_p\}$, for each voxel. The eTofts model parameters θ and aleatoric uncertainty (σ_a) are iteratively updated during the network training to learn a posterior distribution of the parameters and aleatoric uncertainties for θ. A point estimate of the parameters is obtained by sampling from the posterior distribution. In contrast, the epistemic uncertainties $\sigma_e = \{\sigma_e - k_{\mathrm{trans}}, \sigma_e - v_e$ and $\sigma_e - v_p\}$ are not the direct outputs of the BNN. In order to obtain the epistemic uncertainty, the trained network is run n times ($n = 100$) and the standard deviation of the model prediction is computed. This quantifies the average variance of the model prediction. We therefore estimate, for each input $C(t)$ and $C_p(t)$, the mean of the parameters θ, the aleatoric (σ_a) and epistemic (σ_e) uncertainties.

2.2. Extended tofts model

The eTofts model (Tofts 1997) is a two compartmental model, which describes the flow of the contrast agent between the extavascular extracellular space (EES) and the plasma space (vascular space). The eTofts model is solved to estimate three physiological parameters namely, the transfer constant (k_{trans}), fractional volume of the EES (v_e) and fractional volume of the plasma space (v_p). k_{trans} (min^{-1}) is the rate by which the contrast agent diffuses from the plasma space to the EES while v_e and v_p are the EES and plasma space volume per unit tissue volume, respectively. The eTofts model can be written as (Tofts 1997, Heye *et al* 2016):

$$C(t) = v_p C_p(t) + k_{\mathrm{trans}} \int_0^t C_p(t' - \Delta t) e^{-\left(\frac{k_{\mathrm{trans}}}{v_e}\right)(t - t' - \Delta t)} dt' \qquad (1)$$

where $C(t)$ is the time-dependent CA concentration in the tissue, $C_p(t)$ is the concentration of the CA in plasma space over time t and Δt is the time delay of the arrival of the CA at the liver tissue.

2.3. Uncertainty quantification

In the proposed framework, two separate sources of uncertainty are modeled. Aleatoric uncertainty represents the ambiguity that exists in the data. It cannot be reduced by having more training data as it is inherent to the data (Hüllermeier and Waegeman 2021, Tanno et al 2021, Muchen et al 2022). Epistemic uncertainty quantifies the uncertainties in the trained network that emerge from limited training data (Klepaczko et al 2020). This uncertainty describes the presence of features that do not exist in the training data, and can be reduced by adapting the training data (Hüllermeier and Waegeman 2021, Tanno et al 2021, Muchen et al 2022). In the rest of this section, we explain how these two sources of uncertainties are estimated.

2.3.1. *Aleatoric uncertainty and heteroscedastic noise model*

A heteroscedastic noise model (Glang et al 2020), which depends on the input data, is applied to quantify the aleatoric uncertainties. We applied the heteroscedastic noise model because this model describes the variance as a function of the value of the input. For the DCE-data used in this study, the main source of incoherent 'noise-like' artefacts are residual undersampling artefacts. The amplitude of these artefacts is expected to scale with the overall amplitude of the image at each time point. The loss function of the BNN applies the heteroscedastic noise model to optimize the input-dependent varying variance, i.e. the aleatoric uncertainty, as a function of the difference between the reference and estimated parameters. The aleatoric uncertainty is enforced to be large when the difference between the reference and estimated parameters is large (equation (2)). This ensures robustness of the network to noise and outliers in the training data (Tanno et al 2021) and is crucial for a robust DCE parameter estimation in the presence of noise in the training CTCs.

The loss function given in equation (2) uses this noise model to optimize the training of the BNN. During the training process, the parameters of the eTofts model $\hat{\theta}$ and the aleatoric uncertainties (σ_a) of each parameter are both optimized through minimization of this loss function. The loss function is a sum of three terms The first term is the squared sum of differences between the ground truth parameters θ_i and estimated parameters $\hat{\theta}_i$ divided by the aleatoric uncertainty σ_{a_i} for each voxel i. This denominator enforces σ_{a_i} to be large when the difference between the reference and estimated parameters is large i.e. high uncertainty is given to voxels with high error. While the second term prevents the spread of σ_{a_i} from growing indefinitely, the third term $\frac{n}{2}\log(2\pi)$ represents a constant from the Gaussian likelihood function (Glang et al 2020).

$$L = \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\theta_i - \hat{\theta}_i}{\sigma_{a_i}}\right)^2 + \sum_{i=1}^{n}\log(\sigma_{a_i}) + \frac{n}{2}\log(2\pi). \tag{2}$$

2.3.2. *Epistemic uncertainty and bayesian inference*

Given the CTCs, BNN are able to quantify the epistemic uncertainties by providing a posterior distribution of the parameter estimates. In this work, Bayesian inference (Mittermeier et al 2019) is employed for capturing the epistemic uncertainty due to insufficient training data. The exact Bayesian inference of neural network weights is intractable. Therefore, a variational approximation of the true posterior distribution is required. Variational approximation techniques represent weights of neural networks as a probability distribution instead of point estimates. Thus, a variational approximation of the true posterior is given by $q(w|\beta)$, where β denotes the parameters of a Gaussian distribution learned by minimizing the variational free energy shown in equation (3) (Blundell et al 2015).

$$f(C(t), \beta) = \mathrm{argmin}_{\beta}\, KL[q(w|\beta)||p(w)] - E_{q(w|\beta)}[\log p(C(t)|w)] \tag{3}$$

which can be approximated as:

$$f(C(t), \beta) \approx \sum_{j=1}^{m}\log q(w^j|\beta) - \log p(w^j) - \log p(C(t)|w^j). \tag{4}$$

The cost function in equation (3) has two terms, namely, the Kullback-Leibler (KL) divergence and the likelihood function. The first term with the KL divergence is data independent and is used to compute the epistemic uncertainty by learning a posterior distribution of the network weights, $q(w|\beta)$. The prior $p(w)$ provides information about the distribution of the weights prior to observing the CTCs. We have used standard normal Gaussian priors for the weights of the BNN.

During training, we learn the parameters (mean and standard deviation) of the Gaussian distribution $\beta = (\mu, \alpha)$ by minimizing the KL divergence in equation (3). We assumed that β follows a Gaussian distribution. The second term on the other hand determines the data likelihood. In order to compute the likelihood function, Monte Carlo (MC) approximation is performed by sampling network weights w^j from the variational posterior distribution ($w^j \approx q(w^j|\beta)$) as expressed in equation (4) (Blundell et al 2015).

2.3.3. Combined loss function

The combined loss function, shown in equation (5), jointly optimizes the aleatoric and epistemic uncertainties

$$L = \underbrace{\frac{1}{2}\sum_{i=1}^{3}\left(\frac{\theta_i - \hat{\theta}_i}{\sigma_{\theta_i}}\right)^2 + \sum_{i=1}^{3}\log(\sigma_{\theta_i}) + \frac{3}{2}\log(2\pi)}_{a} + \underbrace{\sum_{j=1}^{m}\log q(w^j|\beta) - \log p(w^j)}_{b}. \tag{5}$$

The first term (*a*) is the heteroscedastic Gaussian noise model in equation (2), which is evaluated at the end of the forward pass to capture the aleatoric uncertainty. The second term (*b*) is evaluated layer-wise to reduce the KL divergence between the approximate posterior and prior distributions of the weights to capture the epistemic uncertainty (Blundell *et al* 2015).

3. Experiments

The proposed deep learning framework was trained on simulated data. We evaluated the performance with simulated and patient data.

3.1. Simulated DCE-MRI data

Due to respiratory motion and relatively large field-of-view (FOV) during imaging, high quality 3D DCE-MRI data of the liver is challenging to acquire. In order to overcome the lack of high quality data, we carried out numerical simulations of CTCs to create a training data set. The eTofts model was applied to generate simulated CTCs $C(t)$, using different combinations of pharmacokinetic parameters, $\theta = (k_{trans}, v_e$ and $v_p)$ and an AIF, describing CA uptake in healthy tissue and tumors. For each patient data, $C_p(t)$ was extracted from manually placed ROI on the hepatic artery. For each $C_p(t)$, we simulated 75 000 CTCs with the following parameter ranges: $k_{trans} \in [0.01, 2]$, $v_e \in [0.01, 1]$ and $v_p \in [0.01, 0.3]$ from evenly spaced values in step size of 0.02 to confine them within physiological ranges (Sourbron and Buckley 2011, Ottens *et al* 2022). The temporal resolution of $C(t)$ and $C_p(t)$ was 6 seconds.

Noise level between patients can vary due to variation in patient size, administered contrast agent dose, residual motion and under-sampling artefacts (Jiao *et al* 2020, Ippoliti *et al* 2021, Pandey *et al* 2021). To simulate the effect of signal noise and residual under-sampling artefacts in *in vivo* patient data (Garpebring *et al* 2013, Bliesener *et al* 2020, Klepaczko *et al* 2020, Ottens *et al* 2022), Gaussian noise (n_i) was added to the CTCs, i.e. $C(t)_n = C(t_i) + n_i$. The noise were random samples drawn from a Gaussian distribution with mean ($\mu = 0$), and standard deviation:

$$\lambda = NL{*}C(t)_{\max}. \tag{6}$$

NL was a scaling factor denoting noise levels (NL) in percentages and $C(t)_{\max}$ was the maximum value of the CTCs in noise-free tissues. Overall, the simulated data was used to investigate the potential of the proposed framework for accurate physiological parameter and uncertainty estimation on different scenarios that mimic *in vivo* patient data.

3.2. Network training

The simulated CTCs were split into 80%, 10% and 10% training, testing and validation dataset, respectively. The testing dataset was used to asses the generalization of the model while the validation dataset was used to evaluate the convergence of the training algorithm (Fang *et al* 2021). Furthermore, the network weights were initialized with standard normal Gaussian priors and trained with ADAM optimizer with a learning rate of $1e^{-5}$ and mini batches of size 50. The network was implemented in Python using keras library with Tensorflow backend (Abadi *et al* 2016) and was trained for 200 iterations with a GPU workstation (NVIDIA GeForce RTX 2080). Figure 2 showed the changes in training and validation loss over epochs for a BNN trained with 15% noise level. The gradual decrease in the training loss shows the BNN's ability to learn useful features that maps the input CTCs into physiological parameters (Ulas *et al* 2019).

3.3. Patient DCE-MRI data

Data set from five patients with tumor lesions in the liver were used to assess the performance of the BNN trained with simulated data. DCE data of the liver were acquired using a 3T Biograph mMR hybrid scanner (Siemens Healthcare, Erlangen, Germany) for a duration of 5 minutes after administering a bolus of 0.01 mmol kg^{-1} of hepato-specific contrast agent (gadoxeate disodium). It was a continuous acquisition during free-breathing and motion-corrected image reconstruction was used to obtain DCE images with a temporal resolution of 6 seconds. The acquisition parameters were: TR/TE = 3.3 ms/1.36 ms, flip angle = 12 °, FOV=345 × 345 mm^2, spatial

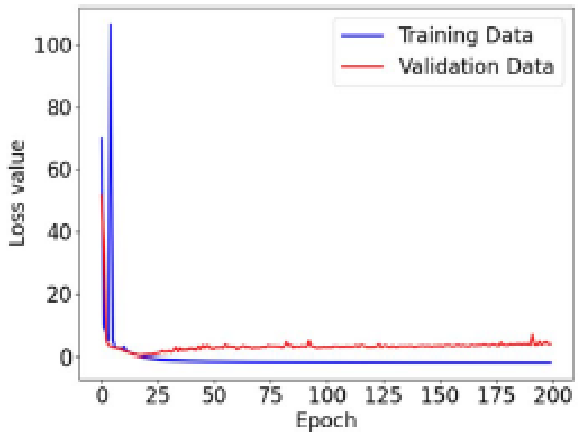


Figure 2. Training and validation loss obtained by training a BNN using CTCs with 15% noise level.

resolution = 1.5 mm^2, and partial Fourier factor = 5/8. More details on the patient data acquisition parameters can be found in (Ippoliti *et al* 2019).

3.4. Evaluation on simulated data

The performance of the proposed framework was evaluated by using simulated data.

3.4.1. Accuracy of parameter estimation

We evaluated the accuracy of parameter estimation using the BNN for different noise levels. λ in equation (6) was set to 0.006, 0.03, 0.06, 0.09 and 0.12, which corresponds to noise levels of 1%, 5%, 10%, 15% and 20%, respectively. Δt in equation (1) was set to zero for the evaluations on the simulated data.

Network performance. We quantitatively evaluated the parameter estimation by calculating the root-mean-squared-errors (RMSE) and R^2 (Inge 1987) values of the output of the network and the reference parameters for different noise levels. In addition, parameter estimation was carried out with a NLLS method for comparison purpose. NLLS fitting was implemented using the inbuilt python minimize function. The NelderMead simplex algorithm (Donald and Lloyd 1975) was applied to minimize the residual between the predicted and measured CTCs voxel-wise. The initial parameters of the NLLS fitting were 0.8, 0.1, 0.01 and 0.001 for k_{trans}, v_e, v_p and Δt, respectively. In addition, we set the lower bounds of the fitting to [0.001, 0.001, 0.001, 0] and upper bounds to [2, 1, 1, 0.4] for k_{trans}, v_e, v_p and Δt, respectively to constrain the parameters to physiologically plausible ranges. Δt was set in minutes.

We applied Akaike information criterion (AIC) (Bozdogan 1987) to compare the model performance between the NLLS fit and the BNN on a test data. The AIC for regression models are calculated as (Banks and Joyner 2017):

$$AIC = n\log(\epsilon) + 2k, \tag{7}$$

where n is the total number of samples, ϵ is the mean square error, and k is the total number of parameters estimated by the model. The best performing model would show a lower AIC.

3.4.2. Uncertainty evaluation of in-distribution (ID) data

We simulated ID training data by applying NL $\in$ [1%, 5%, 10%, 15%, 20%] and evaluated if the aleatoric uncertainty was sensitive to different NL in the data. ID data represented the behavior of the training data. A BNN with NL $\in$ [1%, 5%, 10%, 15%, 20%] was trained. Then, the BNN was applied to test data with a NL of 1%, 5%, 10%, 15% and 20% separately to investigate the effect of different noise levels on the estimated uncertainty. The training and testing noise levels are shown in table 1. The AIF was the same for training and testing.

3.4.3. Uncertainty evaluation of out-of-distribution (OD) data

We evaluated uncertainty estimate for four different OD cases: (i) training data with noise levels different to the application data (OD-NL) (ii) not having enough training data (OD-CTC) (iii) use of inaccurate AIF (OD-AIF) (iv) uncorrected delays in the bolus arrival (OD-Δt). The same AIF was applied for the training and testing data of OD-NL and OD-CTC experiments. For OD-AIF and OD-Δt experiments, different AIF for training and testing were applied.

Table 1. Noise levels and step sizes applied in training and testing of independent BNNs. For all experiments, we have kept the training and testing dataset the same. The difference are the added NL, step sizes and whether the AIF and Δt were ID or OD. The BNN for both the simulated and *in vivo* test data were trained based on simulated data. In addition to the parameter combinations, also the figure numbers are given where the results of the respective evaluations are shown.

Data (Train, Test)	Experiments	Train (NL(%))	Test (NL (%))	Train (step size)	Test(step size)	Figure (#)
Simulated, Simulated	ID	1, 5, 10, 15, 20	1, 5, 10, 15, 20	0.02	0.02	3
	OD-NL	5, 10	15, 20	0.02	0.02	4(*a*), (*e*)
	OD-CTC	5, 10	5, 10	0.04	0.02	4(*b*), (*f*)
	OD-AIF	5, 10	5, 10	0.02	0.02	4(*c*), (*g*)
	OD–Δt	5, 10	5, 10	0.02	0.02	4(*d*), (*h*)
Simulated, *In vivo*	ID	10, 15, 20	—	0.02	—	5, 9
	OD-CTC	10, 15, 20	—	0.04	—	6, 7, 8
	OD-AIF	10, 15, 20	—	0.02	—	6, 7, 8

OD-NL. We investigated the impact of applying test data with OD-NL on the uncertainty estimate. A BNN was trained with NL $\in$ [5%, 10%] and tested with individual NL of 15% and 20% (table 1). The AIF was the same for training and testing.

OD-CTC. We evaluated how varying the size of training data influenced parameter and uncertainty estimation. A BNN was trained with CTCs using parameters k_{trans}, v_e and v_p from evenly spaced values in step size of 0.04, 0.02 and 0.02, respectively. The step size for k_{trans} was increased to 0.04 to decrease the training size by 50%. A second BNN was trained with CTCs generated from k_{trans}, v_e and v_p values with step size of 0.02 for all parameters (i.e. 50% more training data). The AIF and NL was the same for training and testing (table 1).

OD-AIF. We investigated the impact of applying test data with OD-AIF on the uncertainty estimate. The reference AIF was modified to a OD-AIF to mimic errors in the estimation of the uptake of a contrast agent in a patient. We broadened the peak amplitude of the AIF by a factor of two and shifted it by six time points (i.e. 36 seconds) to generate an OD-AIF. In the liver, delays of the AIF up to 20 seconds are most likely considered to be physiological (Miyazaki *et al* 2008, Chouhan *et al* 2016). Hence, a delay of 36 s represents an OD-AIF since larger delay are less likely to be physiological. The modified AIF represented an OD or inaccurate AIF. The NL was the same for training and testing (table 1).

OD-Δt. The sensitivity of the uncertainty estimate to the shifting of the CTCs was investigated by applying test data with OD-Δt. The reference AIF was shifted by two time points (i.e. 12 s) to generate CTCs with OD-Δt. The NL was the same for training and testing (table 1).

3.4.4. *Analysis.*

Coverage evaluation. We evaluated the use of uncertainty as a metric to measure errors of parameter estimates. Errors ($\epsilon_i = |\theta_i - \hat{\theta}_i|$) were computed as the absolute difference between the ground truth parameters (θ_i) and BNN parameter estimates ($\hat{\theta}_i$) for each voxel *i*. The errors were a combination of several sources of uncertainties. Hence, we computed the total uncertainty ($\sigma_{t_i} = \sqrt{\sigma_{a_i}^2 + \sigma_{e_i}^2}$) of each voxel *i* as a combination of the aleatoric (σ_{a_i}) and epistemic uncertainties (σ_{e_i}). The association between the estimated uncertainties and errors of parameter estimates were evaluated by computing the number of voxels (%) with errors below the 90% confidence interval of the total uncertainty, i.e. $\epsilon_i < z^{*}\sigma_{t_i}$, where *z* was the critical value of 1.64.

3.5. *In vivo application*

The BNN trained with simulated data was evaluated by applying it to patients data set to estimate physiological parameters and their uncertainties.

3.5.1. *Parameter estimation*

Parameter estimation was performed for every voxel within the liver. The liver is most likely characterized by delays of the AIF up to 20 seconds, which are considered to be physiological (Miyazaki *et al* 2008, Chouhan *et al* 2016). We applied Δt values up to 18 seconds in steps of one time-point (6 seconds) and generated CTCs with the shifted AIFs. First, a network was trained using a combination of CTCs simulated with NL $\in$ [10%, 15%, 20%], $\Delta t \in$ [0 s, 6 s, 12 s, 18 s] and five AIFs extracted from patient data. Next, physiological parameter maps (k_{trans}, v_e and v_p) were estimated by applying the trained BNN on a DCE-MR patient data set. During training, the BNN learns the shifts in the CTCs arising from the delays in the arrival of the CA. Hence, the inference on patient data will be non-uniform across different voxels of a liver tissue.

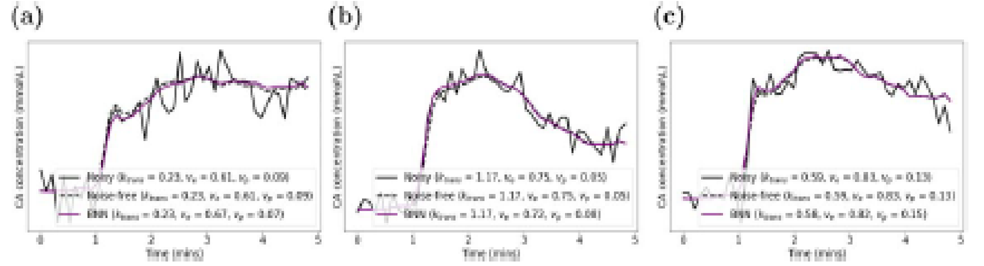


Figure 3. Reference (noise-free and noisy, i.e. 5% noise affected) and BNN CTCs for three different pixels (a)–(c).

3.5.2. *Computational cost*

We evaluated the time cost of the BNN to estimate the eTofts model parameters, aleatoric and epistemic uncertainties from a DCE-MR slice with 192 × 192 voxels on a GPU. In addition, the time cost of the NLLS fitting for estimating the eTofts model parameters was evaluated using the Python time module (Fang *et al* 2021) on a CPU.

3.5.3. *Uncertainty evaluation of ID data*

We simulated ID training data by applying NL ∈ [10%, 15%, 20%], $\Delta t \in$ [0 s, 6 s, 12 s, 18 s] and AIFs from five patients, respectively. Although the true NL is unknown in patient data, we selected the NLs for training visually such that CTCs generated with these NL could be regarded as ID noise levels. The AIFs of the training CTCs were ID as we applied different Δt values that mimic the possible AIF shifts in the *in vivo* data.

3.5.4. *Uncertainty evaluation of OD data*

We evaluated uncertainty estimates for two different OD cases: (i) not having enough training data (OD-CTC) and (ii) use of inaccurate AIF (OD-AIF).

OD-CTC. We simulated an insufficient number of training samples to investigate the influence of lack of CTCs in the training data set on the estimated uncertainty. A BNN was trained with simulated CTCs using parameters k_{trans}, v_e and v_p from evenly spaced values in step size of 0.04, 0.02 and 0.02 and five patients AIF, respectively. This indicates the case of 50% less training data as compared to CTCs generated with step size of 0.02 for all parameters.

OD-AIF. We simulated the use of inaccurate AIFs for generating the training samples and investigated their effect on the parameters and uncertainty estimate. A BNN was trained with CTCs generated without the testing data AIF. The patient data with this OD-AIF for validating the BNN was a unique scan with the same acquisition parameters as the training data as recommended in good machine learning practices (Aggarwal *et al* 2023).

3.5.5. *Regions-of-Interest (ROI) analysis*

We manually chose two ROIs, with one encompassing a tumor lesion and the other healthy liver tissue. The parameter estimates for the healthy and tumor ROI by the BNN and NLLS method were compared. We evaluated the statistical significance of difference of each parameter estimate of both a healthy and a tumor ROI. In addition, the statistical significance of differences of parameter estimates for the tumor and normal tissue ROI was evaluated.

4. Results

4.1. Evaluation of simulated data

4.1.1. *Accuracy of parameter estimation*

Examples of noise-free and 5% noise affected reference CTCs from a test data are shown in figure 3. The predicted curves approximated the reference noise-free curves very well. In addition, BNN determined k_{trans}, v_e and v_p parameter values close to the reference. Table 2 summarized the RMSE and R^2 values for 0%, 1% ,5%, 10% and 15% NLs for the NLLS method and the BNN. NLLS achieved the lowest RMSE and highest R^2 value for k_{trans} for 0% noise level. For all other noise levels and parameters, BNN achieved lower RMSE and higher R^2 values than NLLS. RMSE of BNN was on average of all noise levels by 33% ± 1.9% for k_{trans}, 22% ± 6% for v_e and 89% ± 5% for v_p lower than the NLLS method. In addition, NLLS was more sensitive to noise than BNN with a performance deterioration already for a 1% noise level.

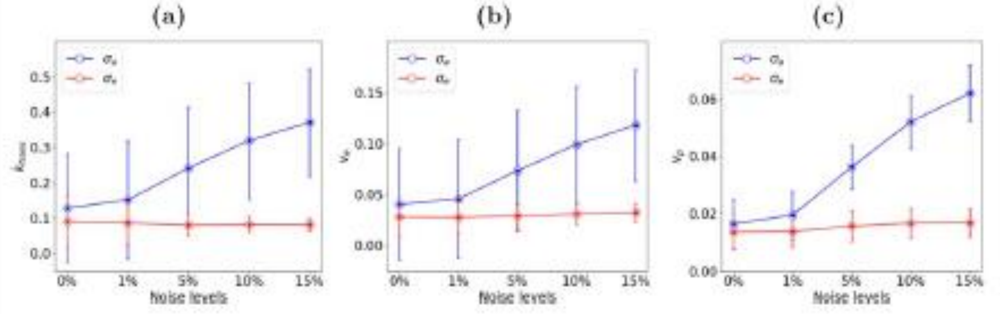


Figure 4. BNN estimates of aleatoric (σ_a) and epistemic (σ_e) uncertainties for varying NLs for (a) k_{trans}, (b) v_e and (c) v_p.

Table 2. RMSE and R^2 between reference and estimated parameters for NLLS and BNN. Best results (lowest RMSE and highest R^2) are shown in bold.

Noise	NLLS						BNN					
	RMSE			R^2			RMSE			R^2		
	k_{trans}	v_e	v_p	k_{trans}	v_e	v_p	k_{trans}	v_e	v_p	k_{trans}	v_e	v_p
0%	**0.10**	0.08	0.27	**0.97**	0.92	0.18	0.16	**0.05**	**0.01**	0.92	**0.97**	**0.99**
1%	0.27	0.09	0.29	0.80	0.90	0.15	**0.19**	**0.06**	**0.01**	**0.89**	**0.96**	**0.98**
5%	0.41	0.1	0.31	0.59	0.88	0.14	**0.27**	**0.08**	**0.03**	**0.77**	**0.92**	**0.88**
10%	0.51	0.12	0.33	0.42	0.83	0.10	**0.34**	**0.10**	**0.05**	**0.65**	**0.88**	**0.70**
15%	0.58	0.15	0.35	0.32	0.76	0.07	**0.38**	**0.12**	**0.06**	**0.55**	**0.83**	**0.56**

The AIC values for the BNN decreased by 28.5% ±7.5%, 12.7% ±6% and 12.2% ±2.7% for k_{trans}, v_e and v_p on average over all noise levels, respectively.

4.1.2. Uncertainty evaluation of ID data

Figure 4 illustrates BNN estimates of aleatoric and epistemic uncertainties for different NLs.

The aleatoric uncertainties of the test data increased ($p < 0.001$) when NL was increased from 5% to 10%, 15% and 20% as presented in figure 4 for k_{trans}, v_e and v_p. The network yielded lower aleatoric uncertainties for lower NLs and higher uncertainties for higher NLs. In addition, testing a BNN with ID data changed the epistemic uncertainty (figure 4) by a smaller value of 0.08 ± 2.7%, 0.03 ± 3.4%, 0.02 ± 5% for k_{trans}, v_e and v_p on average of all noise levels, respectively. Overall, the epistemic uncertainties look similar for all ID test data.

4.1.3. Uncertainty evaluation of OD data

For all OD data, the epistemic uncertainty (figures 5(e), (f), (g) and (h)) increased as expected ($p < 0.0001$) whereas the aleatoric uncertainty increased for OD-CTC (figure 5(b)), OD-AIF (figure 5(c)) and OD-Δt (figure 5(d)). For OD-NL, it decreased significantly ($p < 0.0001$) as shown in figure 5(a).

4.1.4. Analysis

Coverage evaluation. A BNN tested with ID-NLs had a coverage greater than 94% for all NLs and parameters, k_{trans}, v_e and v_p as summarized in table 3. The percentage of voxels with ID-NLs significantly decreased ($p < 0.05$) with increasing noise levels. For a BNN tested with OD-NLs, the coverage decreased for the OD data ($p < 0.001$).

4.2. *In vivo* application

4.2.1. Parameter estimation

An exemplary slice of the DCE-MR data acquired at $t = 4.1$ min showing the anatomy of the liver is displayed in figure 6(a). Physiological parameter maps estimated by the NLLS method and the trained BNN are illustrated in figure 6(b). Tumor lesions showed high k_{trans} and v_p and low v_e values. Generally, BNN showed similar parameter maps with the NLLS giving less noisy estimations, particularly for k_{trans} and v_e. For v_p, BNN and NLLS showed large differences. The estimated parameter maps for the BNN trained with OD training data is shown in figure 7(b). The results for OD-CTCs and OD-AIFs also showed high k_{trans} and v_p and low v_e values for tumor lesions.

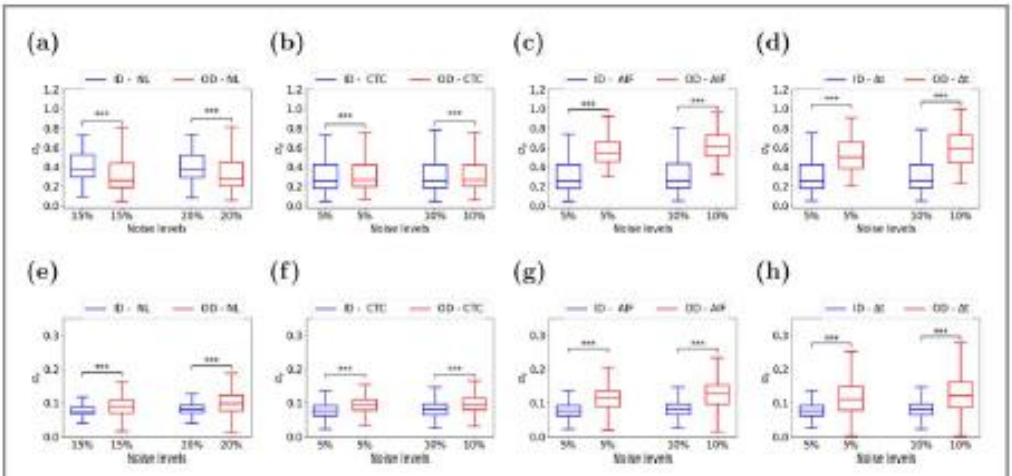


Figure 5. Aleatoric (σ_a) and epistemic(σ_e) uncertainties of k_{trans} for OD-NL (a), (e), OD-CTC(b), (f), OD-AIF (c), (g) and OD-Δt(d), (h), respectively. '****' denotes a statistical significance level of $p < 0.0001$.

Table 3. Percentages of voxels with error estimates within the 90% confidence interval of the total uncertainty. The coverage was greater for ID-NL (shown in bold) as compared to OD-NL.

Noise level (NL)	ID-NL			OD-NL		
	k_{trans}	v_e	v_p	k_{trans}	v_e	v_p
0%	99	99	99	—	—	—
5%	99	98	99	—	—	—
10%	97	97	98	—	—	—
15%	**95**	**96**	**96**	88	92	89
20%	**94**	**94**	**94**	84	89	86

Moreover, the DCE-MR slice of four more additional patients are illustrated in figures 10(a)–(d). The corresponding k_{trans} maps estimated by the BNN for the four patients are shown in figures 10(e)–(h). The tumors showed high k_{trans} values.

4.2.2. *Computational cost*

The BNN required 0.1 min to estimate the eTofts model parameters, aleatoric and epistemic uncertainties from a DCE-MR slice with 192 × 192 voxels on a GPU. In contrast, the NLLS fitting required 270 min to estimate eTofts model parameters on a CPU. The BNN substantially decreased the time cost and was faster than the NLLS fitting.

4.2.3. *Uncertainty evaluation of ID data*

The aleatoric and epistemic uncertainties of the DCE-MR slice in figure 6(a) are separately shown in figure 8 and figure 9, respectively. In addition, the aleatoric and epistemic uncertainties of the four patients in figures 10(a)–(d) are shown in figures 10(i)–(l) and figures 10(m)–(p), respectively. Similar to the aleatoric and epistemic uncertainties of the ID data of figures 8 and 9, the uncertainties for the four patients were high, especially in the regions of the tumor.

4.2.4. *Uncertainty evaluation of OD data*

Aleatoric uncertainties of parameters estimated by a BNN trained with smaller training data size (OD-CTC) showed similar values to the results of the ID data as shown in figure 8. On the other hand, the aleatoric uncertainties for k_{trans} and v_p of the OD-AIFs showed increased uncertainty than the ID BNN.

However, epistemic uncertainties of all parameters, particularly k_{trans} and v_p increased for the OD data (figure 9).

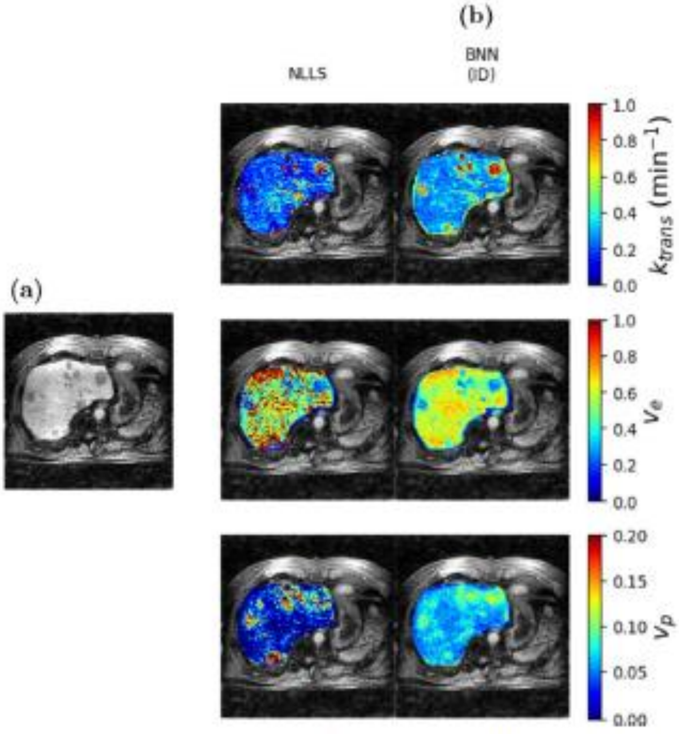


Figure 6. (a) Exemplary slice of the DCE-MR data for a patient with hepatic tumor lesions (depicted as 'darker' spots in the liver) acquired at $t = 4.1$ min, (b) Estimated parameter maps using NLLS and BNN for k_{trans}, v_e and v_p. Each row represents a given parameter and the two columns for each row represent the NLLS and BNN.

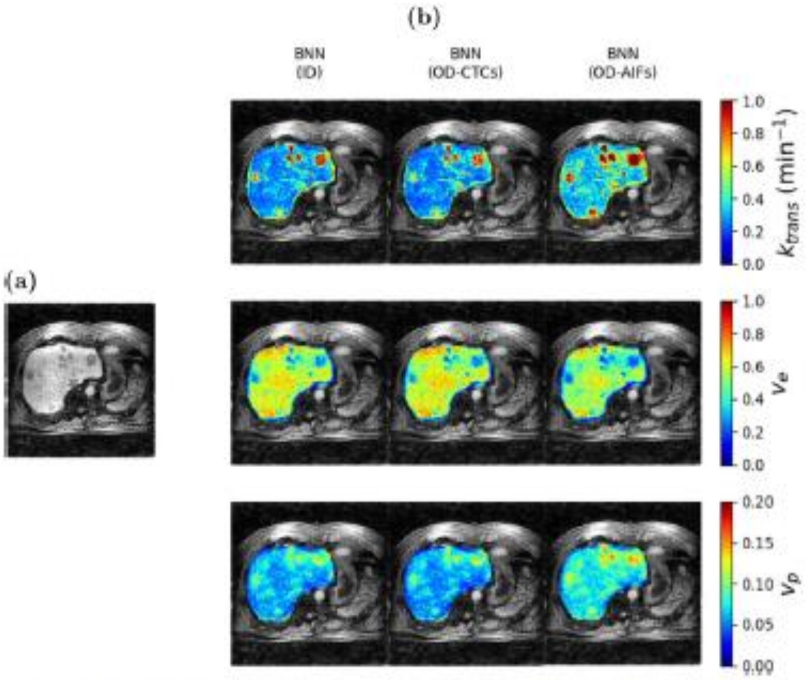


Figure 7. (a) Exemplary slice of the DCE-MR data for a patient with hepatic tumor lesions (depicted as 'darker' spots in the liver) acquired at $t = 4.1$ min, (b) estimated parameter maps (k_{trans}, v_e and v_p) using a test data set consisting of ID data, OD-CTCs and OD-AIFs. Each row represents a given parameter and the first two columns for each row represent the ID and OD-CTCs, while the third column presents the OD-AIFs.

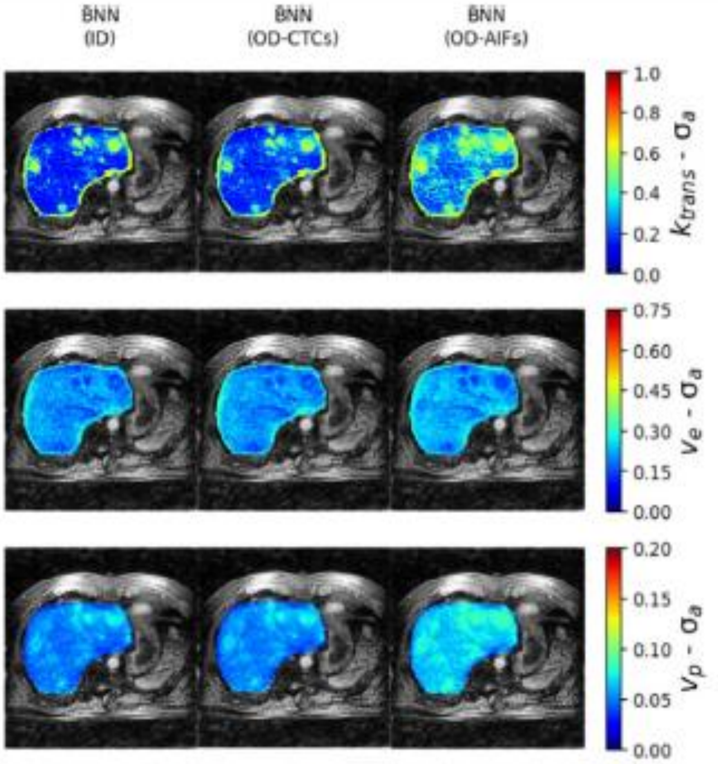


Figure 8. Aleatoric uncertainties (σ_a) estimated by BNN for ID data, OD-CTC and OD-AIF for k_{trans}, v_e and v_p. Each row represents a given parameter and the first two columns for each row represent the ID and OD-CTCs, while the third column presents the OD-AIFs.

4.2.5. Regions-of-interest (ROI) analysis

The ROIs for a tumor lesion and a healthy region are shown in figure 11(a). For the NLLS (figure 11(b)), k_{trans} and v_e for the tumor showed an overlap. In addition, k_{trans} and v_p of the tumor overlapped with k_{trans} and v_p of the healthy region. High variance for the parameter estimates was observed in k_{trans} for the tumor and v_e of the healthy region. On the contrary, BNN parameter estimates for k_{trans}, v_e and v_p in the selected ROIs provided a clear distinction and discrimination between healthy and tumor regions as shown in figure 11(c). Differences between healthy and tumor ROIs were significant for all parameters ($p < 0.0001$). In addition, the difference between each parameter estimate of both a healthy and a tumor ROI were significant ($p < 0.0001$). The difference between parameter estimates (k_{trans}, v_e and v_p) for a healthy and tumor ROI were also significant ($p < 0.0001$). Tumors were characterized by high k_{trans} and v_p and low v_e values in both BNN and NLLS.

5. Discussion

We implemented a deep learning based framework for both parameter and uncertainty estimation using simulated data, and showed its application in patients DCE-MRI data. Our proposed framework improved the performance of parameter estimation and was more robust to noise than the NLLS method as summarized in table 2. Moreover, BNN was able to assign each voxel its own aleatoric and epistemic uncertainty instead of assigning global uncertainty values for certain set of samples. Such voxel-based uncertainties provide more reliable measurements as compared to global bounds that are prone to either over- or under-estimations (Bliesener *et al* 2020).

BNN was sensitive to different NLs if all the NLs were ID of the training data (figure 4). Such sensitivity to variations in NLs is important when dealing with patient data because the NLs in DCE-MR varies due to variation in patient size, administered contrast agent dose, and residual motion and under-sampling artefacts (Jiao *et al* 2020, Ippoliti *et al* 2021, Pandey *et al* 2021). Therefore, the BNN could be applied to different NLs and it could capture the NLs in the data.

As expected, for ID data different NLs had only a very small effect on the epistemic uncertainty with an average change of 0.08 ± 2.7%, 0.03 ± 3.4%, 0.02 ± 5%, for k_{trans}, v_e and v_p over all noise levels, respectively.

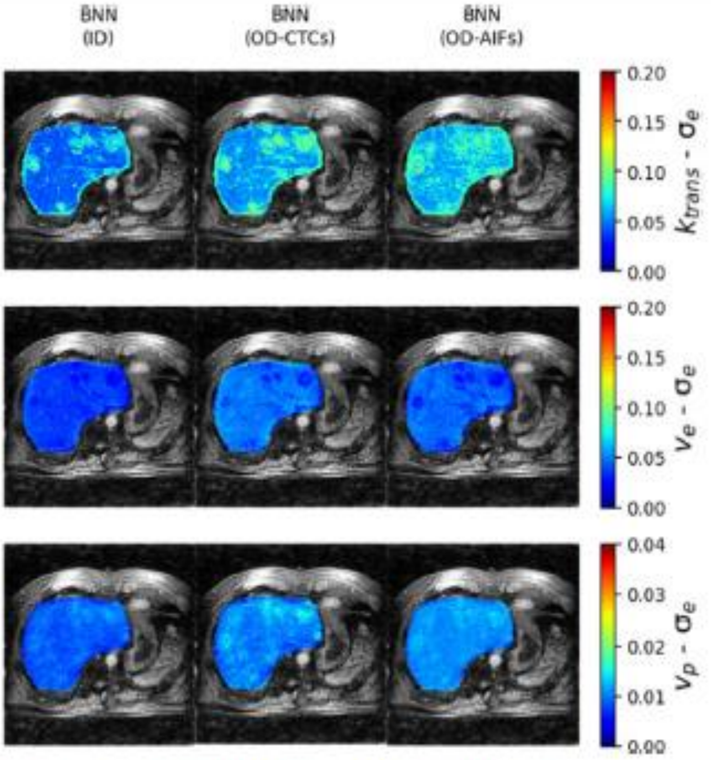


Figure 9. Epistemic uncertainties (σ_e) estimated by BNN for ID data, OD-CTC and OD-AIF for k_{trans}, v_e and v_p. Each row represents a given parameter and the first two columns for each row represent the ID and OD-CTCs, while the third column presents the OD-AIFs.

The epistemic uncertainty increased when a BNN was evaluated with OD data such as, NLs, CTCs, AIF and Δt as illustrated in figures 5(e), (f), (g) and (h), respectively. Insufficient training data increased the variance of the posterior distribution, thereby resulting in higher epistemic uncertainty (Tanno *et al* 2021).

BNN demonstrated higher aleatoric uncertainty for k_{trans} with an out-of-distribution test data AIF and Δt as depicted in figures 5(c) and (d), respectively. This suggests that OD data also impacts on the estimation of the aleatoirc uncertainty. The network yields less reliable output for OD data which is correctly reflected in a higher epistemic uncertainty. In addition, the network cannot estimate the aleatoric uncertainty as well as for ID data leading to an apparent increase of the aleatoric uncertainty. Further studies are required to investigate this effect in more detail.

Overall, our proposed BNN was able to separate the different sources of uncertainty which resulted from the underlying noise or residual under-sampling artefacts (aleatoric uncertainty) or insufficient training data (epistemic uncertainty).

The RMSE was associated with the total uncertainties and the coverage was greater than 90% for all ID-NL and greater than 80% for OD-NL experiments as shown in table 3. The association between the errors and uncertainty estimates were better characterized with ID-NL because a BNN could capture the variations of NL if the testing data was ID of the training data. This indicates the use of uncertainty as a metric to measure errors of parameter estimates.

Experiments on five patients data set revealed the potential of BNN trained with a simulated data set in estimating k_{trans}, v_e and v_p of eTofts model parameters (figures 6(b) and 10). The results proved that the simulated data was a good representation of the patient data and the BNN was well trained and generalized the training data (Zou *et al* 2020). Consistent to the simulations, k_{trans} of a patient data set showed the highest aleatoric uncertainty followed by v_p. For the epistemic uncertainty, v_e showed the lowest values. This was also consistent with the results from the simulation.

The analysis with and without incorporating the test data AIF showed the utility of epistemic uncertainty in explaining performance of parameter estimation with respect to insufficient training data. The epistemic uncertainty of the parameters increased when the test data AIF was not taken into account. Moreover, the epistemic uncertainty increased when a larger step size was applied. This demonstrates the feasibility of epistemic

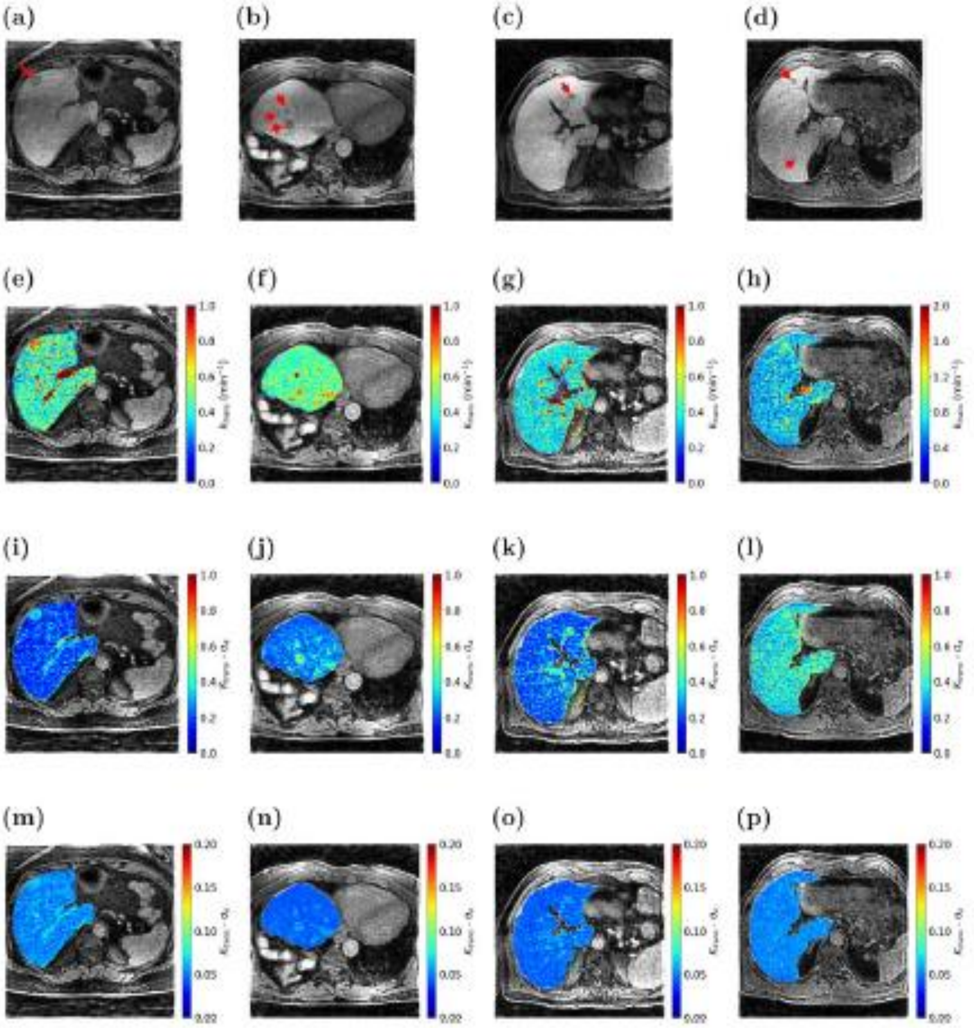


Figure 10. (a)–(d) Exemplary slices of the DCE-MR data for four patients with hepatic tumor lesions (depicted as 'darker' spots in the liver), (e)–(h), estimated k_{trans} maps for each patient, (i)–(l), aleatoric uncertainties (σ_a) for k_{trans}, (m)–(p), epistemic uncertainties (σ_e) for k_{trans}. Each column represent the corresponding estimates for a patient.

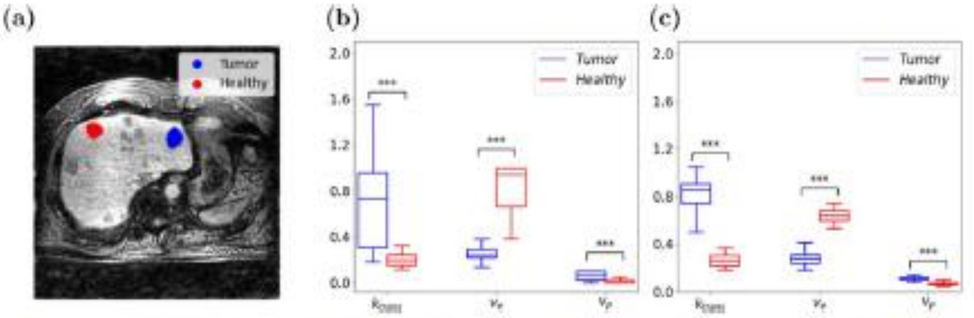


Figure 11. Comparison of estimated parameters in healthy and tumor regions for NLLS and BNN (a) ROIs, (b) NLLS and (c) BNN. '***' denotes a statistical significance level of $p < 0.0001$.

uncertainty in explaining insufficient training data. On the other hand, aleatoric uncertainties were not affected by lack of training data and could not be decreased with more training data (Bliesener *et al* 2020, Tanno *et al* 2021).

Parameter estimates for a tumor and a healthy ROI of a patient data set (figure 11(c)) indicated BNN's ability to discriminate between healthy and tumor regions. Due to the cost functions convergence to local minimas and dependence on parameter initialization (Bliesener et al 2020, Ottens et al 2022, Rudolf et al 2022), parameter estimation of the NLLS of the patient data showed noisy predictions (figure 6(b)) and high variance (figure 11(b)). The lower AIC values of the BNN for all noise levels showed its superior performance as compared to the NLLS fit.

In this study, we used a single input two-compartment eTofts model. The use of a dual input two-compartment eTofts model (Yang et al 2016, Chouhan et al 2017, Liu et al 2019, Li et al 2020) which takes the arterial input function (AIF) and portal input function (PIF) as an input would also be possible with our network design and could lead to a more accurate estimation of perfusion in the liver. We applied Δt values up to 18 s in steps of one time-point (6 s) and generated CTCs with the shifted AIFs to confine it with physiological ranges of the liver (Miyazaki et al 2008, Chouhan et al 2016). Estimation of Δt as an output parameter would also be possible with our network design and could lead to a more accurate estimation of perfusion in the liver. Future work with the estimation of Δt as an output parameter could enhance the reproducibility of liver DCE MRI-based physiological parameters estimation and could be of clinical interest (Chouhan et al 2016). Further studies are required to investigate the usability of Δt as a diagnostic parameter. The evaluations with five patients showed the ability of the BNN to estimate physiological parameters better than the NLLS. Nevertheless, larger studies are required to carry out a full assessment of the benefits and limitations of the proposed approach in clinical practice (Aggarwal et al 2023). One limitation of the BNN is the use of a fixed temporal resolution of 6 seconds for simulating the training $C_p(t)$ and $C(t)$. Future work is required to improve the generalizability of the BNN to consider varying temporal resolutions.

The epistemic uncertainties indicated how well the BNN performed with respect to less training data and cases where the training data differs from the application data. One limitation of this work is, that it does not utilise XAI methods. The use of explainable AI (XAI) techniques would give further rationales on the estimated parameters of the BNN. One possibility could be the use of post-hoc XAI techniques, such as feature attribution methods (Jiménez-Luna et al 2020, Letzgus et al 2022) to investigate the trained BNN parameter estimates by applying input CTCs simulated with OD-AIF or OD-Δt. This can be used to visually show the BNN's performance with OD-data. In addition, the resilience of the trained BNN to noise in estimating the physiological parameters could be visualized with feature attribution methods. By varying the noise levels in the input CTCs, the effect of noise in the input on the trained BNN parameter estimates can be visualized. Future work with the application of the above mentioned XAI methods could further improve the interpretability of the BNN.

6. Conclusion

In this work, we were able to estimate aleatoric and epistemic uncertainties associated with parameter estimation of DCE-MR data. Our proposed framework was evaluated by numerical simulations and applied to data set of patients suffering from hepatic tumor lesions. The proposed BNN outperformed the standard NLLS method in accuracy of parameter estimation and computational cost. In addition, BNN was able to estimate the different sources of uncertainty which resulted from the underlying noise or residual under-sampling artefacts (aleatoric uncertainty) or insufficient training data (epistemic uncertainty). Aleatoric uncertainties increase with increasing NLs while, the epistemic uncertainties increase for AIFs, NLs, Δt or CTCs, which were not part of the training data (out-of-distribution).

Acknowledgments

The authors gratefully acknowledge funding from the German Research Foundation (GRK2260, BIOQIC).

Data availability statement

The data cannot be made publicly available upon publication due to legal restrictions preventing unrestricted public distribution. The data that support the findings of this study are available upon reasonable request from the authors.

Ethical statement

This work included a patient data who provided a written informed consent and the study was prospectively approved by and registered with Charité Ethics Committee (Trial Registration Number: EA4/051/17). The research was conducted in accordance with the principles embodied in the Declaration of Helsinki and in accordance with local statutory requirements.

References

Abadi M *et al* 2016 Tensorflow: large-scale machine learning on heterogeneous distributed systems *Proc. of the 12th USENIX Conf. on Operating Systems Design and Implementation, OSDI'16, USENIX Association* pp 265–83

Aggarwal K, Marina M J, Keerthi S R, Gilberto G and Geethanath S 2023 Developing and deploying deep learning models in brain MRI: a review *NMR in Biomedicine* e5014

Ahearn T S, Staff R T, Redpath T W and Semple S I 2005 The use of the Levenberg–Marquardt curve-fitting algorithm in pharmacokinetic modelling of DCE-MRI data *Phys. Med. Biol.* 50 85–92

Banks H T and Joyner M L 2017 AIC under the framework of least squares estimation *Appl. Math. Lett.* 74 33–45

Berks M, Little R A, Watson Y, Cheung S, Datta A, O'Connor J P B, Scaramuzza D and Parker G J M 2021 A model selection framework to quantify microvascular liver function in gadoxetate-enhanced MRI: application to healthy liver, diseased tissue, and hepatocellular carcinoma *Magn. Reson. Med.* 86 1829–44

Bliesener Y, Acharya J and Nayak K S 2020 Efficient DCE-MRI parameter and uncertainty estimation using a neural network *IEEE Trans. Med. Imaging* 39 1712–23

Blundell C, Cornebise J, Kavukcuoglu K and Wierstra D 2015 Weight uncertainty in neural network *Proc. of the 32nd Int. Conf. on Machine Learning, vol 37 of ICML'15, PMLR* pp 1613–22

Bozdogan H 1987 Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions *Psychometrika* 52 345–70

Chouhan M D, Bainbridge A, Atkinson D, Punwani S, Mookerjee R P, Lythgoe M F and Taylor S A 2016 Estimation of contrast agent bolus arrival delays for improved reproducibility of liver DCE MRI *Phys. Med. Biol.* 61 6905–18

Chouhan M D, Bainbridge A, Atkinson D, Punwani S, Mookerjee R P, Lythgoe M F and Taylor S A 2017 Improved hepatic arterial fraction estimation using cardiac output correction of arterial input functions for liver DCE MRI *Phys. Med. Biol.* 62 1533–46

Choyke P L, Dwyer A J and Knopp M V 2003 Functional tumor imaging with dynamic contrast-enhanced magnetic resonance imaging *J. Magn. Reson. Imaging* 17 509–20

Cristina L 2015 Simulating the effect of input errors on the accuracy of Tofts' pharmacokinetic model parameters *Magn. Reson. Imaging* 33 222–35

Cuenod C A and Balvay D 2013 Perfusion and vascular permeability: Basic concepts and measurement in DCE-CT and DCE-MRI *Diagn. Interventional Imaging* 94 1187–204

Donald M O and Lloyd S N 1975 The nelder-mead simplex procedure for function minimization *Technometrics* 17 45–51

Dündar T T, Cetinkaya E, Yurtsever İ, Uysal Ö and Aralaşmak A 2022 Follow-up of high-grade glial tumor; differentiation of posttreatment enhancement and tumoral enhancement by dce-mr perfusion *Contrast Media Mol. Imaging* 2022 6948422

Fang K *et al* 2021 Convolutional neural network for accelerating the computation of the extended Tofts model in dynamic contrast-enhanced magnetic resonance imaging *J. Magn. Reson. Imaging* 53 1898–910

Garpebring A, Brynolfsson P, Yu J, Yu J, Wirestam R, Johansson A, Asklund T and Karlsson M 2013 Uncertainty estimation in dynamic contrast-enhanced MRI *Magn. Reson. Med.* 69 992–1002

Glang F, Deshmane A, Prokudin S, Martin F, Herz K, Lindig T, Bender B, Scheffler K and Zaiss M 2020 Deep CEST 3T: robust MRI parameter determination and uncertainty quantification with neural networks-application to CEST imaging of the human brain at 3T *Magn. Reson. Med.* 84 450–66

Gurney-Champion O J, Mahmood F, van Schie M, Julian R, George B, Philippens M E P, van der Heide U A, Thorwarth D and Redalen K R 2020 Quantitative imaging for radiotherapy purposes. Radiotherapy and oncology *J. Eur. Soc. Therapeutic Radiol. Oncol.* 146 66–75

Hansen M B, Tietze A, Haack S, Kallehauge J, Mikkelsen I K, Østergaard L and Mouridsen K 2019 Robust estimation of hemo-dynamic parameters in traditional DCE-MRI models *PLoS One* 14 e0209891

Heye A K, Thrippleton M J, Armitage P A, Valdés Hernández M, Makin S D, Glatz A, Sakka E and Wardlaw J M 2016 Tracer kinetic modelling for DCE-MRI quantification of subtle blood-brain barrier permeability *NeuroImage* 125 446–55

Hüllermeier E and Waegeman W 2021 Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods *Mach. Learn.* 110 457–506

Inge S H 1987 On the interpretation and use of R^2 in regression analysis *Biometrics* 43 61–69

Ippoliti M, Lukas M, Brenner W, Schaeffter T, Makowski M R and Kolbitsch C 2019 3d nonrigid motion correction for quantitative assessment of hepatic lesions in DCE-MRI *Magn. Reson. Med.* 82 1753–66

Ippoliti M, Lukas M, Brenner W, Schatka I, Furth C, Schaeffter T and Kolbitsch C 2021 Respiratory motion correction for enhanced quantification of hepatic lesions in simultaneous PET and DCE-MR imaging *Phys. Med. Biol.* 66 95012

Jackson A, O'Connor J P, Parker G J and Jayson G C 2007 Imaging tumor vascular heterogeneity and angiogenesis using dynamic contrast-enhanced magnetic resonance imaging *J. Am. Assoc. Cancer Res.* 13 3449–59

Jiao H, Jiang X, Pang Z, Lin X, Huang Y and Li L 2020 Deep convolutional neural networks-based automatic breast segmentation and mass detection in DCE-MRI *Comput. Math. Methods Med.* 2020 2413706

Jiménez-Luna J, Grisoni F and Schneider G 2020 Drug discovery with explainable artificial intelligence *Nat. Mach. Intell.* 2 573–84

Klepaczko A, Strzelecki M, Kociołek M, Eikefjord E and Lundervold A 2020 A multi-layer perceptron network for perfusion parameter estimation in DCE-MRI studies of the healthy kidney *Appl. Sci.* 10 5525

Letzgus S, Wagner P, Lederer J, Samek W, Müller K R and Montavon G 2022 Toward explainable artificial intelligence for regression models: a methodological perspective *IEEE Signal Process Mag.* 39 40–58

Li J, Xue F, Xu X, Wang Q and Zhang X 2020 Dynamic contrast-enhanced mri differentiates hepatocellular carcinoma from hepatic metastasis of rectal cancer by extracting pharmacokinetic parameters and radiomic features *Exp. Therapeutic Med.* 20 3643–52

Liu Q, Gao Y, Wang Y, Du J, Yin Q and Shi K 2019 Diagnostic value of hepatic artery perfusion fraction combined with tgf-β in patients with hepatocellular carcinoma *Oncol. Lett.* 17 5635–41

Mittermeier A, Ertl-Wagner B, Ricke J, Dietrich O and Ingrisch M 2019 Bayesian pharmacokinetic modeling of dynamic contrast-enhanced magnetic resonance imaging: validation and application *Phys. Med. Biol.* **64** 18NT02

Miyazaki S, Murase K, Yoshikawa T, Morimoto S, Ohno Y and Sugimura K 2008 A quantitative method for estimating hepatic blood flow using a dual-input single-compartment model *Br. J. Radiol.* **81** 790–800

Muchen W, Tangfan X, Jiangtao C and Yu L 2022 Differentiating effects of input aleatory and epistemic uncertainties on system output: a separating sensitivity analysis approach *Mech. Syst. Sig. Process.* **181** 109421

Ottens T, Barbieri S, Orton M R, Klaassen R, van Laarhoven H W M, Crezee H, Nederveen A J, Zhen X and Gurney-Champion O J 2022 Deep learning DCE-MRI parameter estimation: application in pancreatic cancer *Med. Image Anal.* **80** 102512

Pandey D, Wang H, Yin X, Wang K, Zhang Y and Shen J 2021 Automatic breast lesion segmentation using continuous max-flow algorithm in phase preserved DCE-MRIs *Health Information Science: 10th Int. Conf.* vol 13079 (Springer) pp 124–37

Rudolf L M, van H, Amedeo C, Marcel B, Mitko V and Cian M S 2022 Physics-informed neural networks for myocardial perfusion MRI quantification *Med. Image Anal.* **78** 102399

Sourbron S P and Buckley D L 2011 On the scope and interpretation of the Tofts models for DCE-MRI. Magnetic resonance in medicine *Magn. Reson. Med.* **66** 735–45

Sourbron S P and Buckley D L 2013 Classic models for dynamic contrast-enhanced MRI *NMR Biomed.* **26** 1004–27

Tanno R, Worrall D E, Kaden E, Ghosh A, Grussu F, Bizzi A, Sotiropoulos S N, Criminisi A and Alexander D C 2021 Uncertainty modelling in deep learning for safer neuroimage enhancement: demonstration in diffusion MRI *NeuroImage* **225** 117366

Tofts P S 1997 Modeling tracer kinetics in dynamic Gd-DTPA MR imaging *J. Magn. Reson. Imaging* **7** 91–101

Tofts P S *et al* 1999 Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols *J. Magn. Reson. Imaging* **10** 223–32

Ulas C *et al* 2019 Convolutional neural networks for direct inference of pharmacokinetic parameters: application to stroke dynamic contrast-enhanced MRI *Front. Neurol.* **9** 1147

Wang P N, Velikina J V, Bancroft L, Samsonov A A, Kelcz F, Strigel v R M and Holmes J H 2022 The influence of data-driven compressed sensing reconstruction on quantitative pharmacokinetic analysis in breast DCE MRI *Tomography (Ann Arbor, Mich.)* **8** 1552–69

Yang J F, Zhao Z H, Zhang Y, Zhao L, Yang L M, Zhang M M, Wang B Y, Wang T and Lu B C 2016 Dual-input two-compartment pharmacokinetic model of dynamic contrast-enhanced magnetic resonance imaging in hepatocellular carcinoma *World J. Gastroenterol.* **22** 3652–62

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 *Med. Phys.* **47** 3447–57

Curriculum Vitae

My curriculum vitae does not appear in the electronic version of my paper for reasons of data protection.

My curriculum vitae does not appear in the electronic version of my paper for reasons of data protection.

My curriculum vitae does not appear in the electronic version of my paper for reasons of data protection.

Publication list

Dejene EM, Brenner W, Makowski MR, Kolbitsch C. Unified Bayesian network for uncertainty quantification of physiological parameters in dynamic contrast enhanced (DCE) MRI of the liver. Phys Med Biol. 2023;68(21)

Dejene EM, Kolbitsch C. Deep learning for quantitative evaluation of motion correction in dynamic contrast enhanced (DCE) MRI. Proceedings of the 32nd Annual Meeting of ISMRM. Toronto, ON, Canada. May 2023.

Dejene EM, Brenner W, Makowski MR, Kolbitsch C. Deep learning for accurate estimation of contrast agent bolus arrival delays in DCE MRI of the liver. Proceedings of the 25th Annual Meeting of DS-ISMRM. Berlin, Germany. September 2023.

Acknowledgments

First and foremost, my honour and appreciation goes to God who stood by me, gave me sufficient grace to overcome all challenges. Thank you, Lord, for the working grace that made everything possible for me.

I would like to give my heartfelt and sincere gratitude to my supervisor, Dr. Christoph Kolbitsch. I am grateful for your advice, continuous guidance, and feedback throughout my PhD Dissertation Research Project. I would like to respectfully acknowledge Prof. Dr. Winfried Brenner and Prof. Dr. Marcus Makowski, for the critical comments and feedback, to my research work. Also, I am thankful to both of you, Prof. Tobias Schäffter and Prof. Ingolf Sack, for your valuable feedback to my overall work.

I would like to acknowledge the funding from the German Research Foundation (GRK2260, BIOQIC). I would like to greatly acknowledge Dr. Judith Bergs, who provided me with invaluable support with regard to the PhD Dissertation process, including pertinent administrative issues during my stay in Berlin, Germany.

I would like to thank with gratitude everyone in Christ Evangelical Church (CEC) in Munich, Germany. My special thanks goes to Pastor Amde Aklilu, Senayit Woldemariam, Selamawit Yibra, Tamirat Abera, Yohanna Yitbarek, Daniel Berhanu, Behailu Getahun, Aman Gashaw, Worku Alemu, Hawi Geleta, Tolosa Hunde, Elsabeth Kassa, Tigist Negash, Deres Gebreeyesus, Emebet Markos, Yeshi Bereda and, Atakelt Sebhatu for the friendship and support in prayers during my stay in Munich. Unforgettable memory takes me back to the late Hanna Abera, whose kindness and support would always be remembered. Overall, I am very thankful for everyone in unison at CEC Munich.

Furthermore, I would like to thank everyone in Berlin Bethlehem Church, especially Pastor Seife Bekele, Misrak Kushe and Eyerusalem Shebrou. My special thanks goes to you Meron Solomon (Merye), for those pleasant and unforgettable times in Berlin. The worship time every Saturday was and is still an energy, impetus for me for the week.

My deep appreciation goes to my father, Prof. Dr. Mashilla Dejene. Dad, thank you for being a great-hearted father, a teacher who shaped my moral and ethical standards and a man who loves my mother so much. Thank you for your continuous encouragement, which has been a great source of inspiration. I would like to acknowledge my brother Eskinder Mashilla for his support and care at all times.

My special and invaluable thanks go to my mother, Alemtsehay Berihun (Ema). Ema, your prayer was and is still continuous; your care for me is selfless; your pieces of advice are full of wisdom and understanding. You have shown me the power of prayer, the meaning of faith, and showed Christ to me. Thank you, Ema, for your friendship, for always being there to guide me, comfort me, lead me to the big plans of God over my life. I would like to extend my special thanks to my husband, Tewodros Awgichew. I really appreciate your support and understanding in all the times I have been busy with my doctoral research work.

Last but not least, my special thanks go to my precious daughter Tsiyon Tewodros. Sometimes the days get longer because I cannot wait to go home and be with you. Tsinu/Li-elt (Joy), may you grow with the wisdom and love of the Lord each and every day. Thank you for bringing more joy to my life.