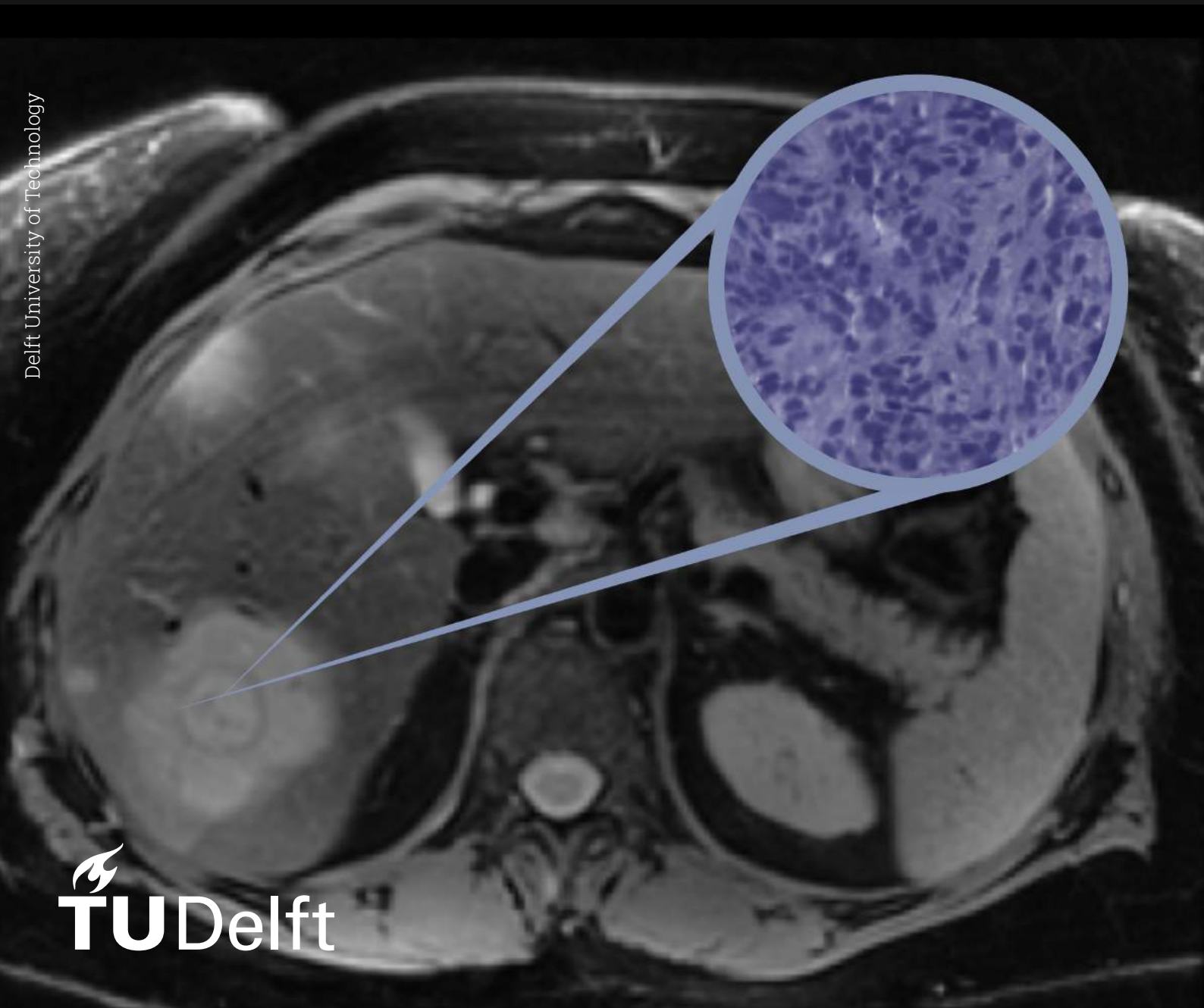
Unraveling Tumour Microstructure through Diffusion MRI using Histology-Powered Artificial Intelligence

Ella Fokkinga



Delft University of Technology

TU Delft

Unraveling Tumour Microstructure through Diffusion MRI using Histology-Powered Artificial Intelligence

by

Ella Fokkinga

to obtain the degree of **Master of Science**
in Biomedical Engineering, Medical Physics
at the Delft University of Technology,
to be defended publicly on Monday the 19th of June at 14:00.

Student number:	4555759
Project duration:	September, 2022 – June, 2023
Daily supervisor:	F. Grussu
Supervisors:	J.A. Hernández Tamames, R. Perez-Lopez
Thesis committee:	F. Vos, D. Poot

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TU Delft

Abstract

Introduction Diffusion-Weighted Magnetic Resonance Imaging (DW-MRI) is a quantitative MRI method, where the diffusion of water molecules is exploited as a source of contrast among tissues. By sensitising DW-MR-measurements to diffusion, we enable the non-invasive quantification of tissue microstructure. The signal observed in a voxel arises from contributions of various diffusion processes within different cellular components (i.e. the tissue microstructure). Many methods exist aiming to disentangle these signal contributions and to compute biologically meaningful microstructure indices. A new, innovative method is using Artificial Intelligence (AI) to directly predict microstructure from DW-MR-signal data, instead of relying on signal representations or biophysical models.

Objective The objective of this project is the development of a novel AI-based method to estimate microstructural parameters in the liver, specifically cell size and intracellular fraction, given paired examples of DW-MRI signals and microstructural features as seen on histology. We aim to outperform the estimation of microstructure by a well-established biophysical model. The method will be developed on pre-clinical data, and will be tested by deployment on *in vivo* patient data.

Methods Seven algorithms were implemented in Python, of which four voxel-wise Machine Learning (ML) (Linear Regression, Piecewise Linear Regression, K-Nearest Neighbour, Random-Forest Regressor (RFR)), one patch-wise ML (PW-RFR), one voxel-wise Deep Learning (DL) (Multi-Layer-Perceptron (MLP)) and one patch-wise DL (PW-MLP). The algorithms were trained in a supervised manner, using MRI-derived data co-registered to histology-derived parametric maps of the targets (i.e. the ground truth) of seven mouse livers. The performance of the algorithms based on three input feature types is analysed: (i) the raw DW-MR-data, (ii) signal representations, (iii) parametric maps computed by fitting a multi-exponential biophysical model. Quantitative evaluation metrics (Mean Absolute Error (MAE) and Relative MAE) and visual inspection are used for performance assessment. The most promising pre-trained algorithm from the pre-clinical study is consequently tested on clinical data of three patients.

Results Based on quantitive evaluation metrics and visual inspection of the predictions in comparison with the histology-derived ground truth, the PW-RFR was found to be the most promising algorithm. With the PW-MLP exhibiting the second best results, the importance of incorporating spatial information, here via the patch-wise approach, is supported. No large differences were observed in the performance of the best algorithm, the PW-RFR, for the three input feature types. From the clinical study appeared that the PW-RFR is able to at least partly capture notable contrasts in the lesion microstructure, and the predictions are in a realistic range in comparison with histology-derived microstructure values as obtained via biopsy of the lesions.

Conclusion The PW-RFR algorithm demonstrates a promising performance in both the pre-clinical and the clinical experiment. This points towards the feasibility of developing tools that learn how to predict histopathological characteristics from DW-MRI on pre-clinical MRI-histology data, and consequently deploy these tools in a clinical setting. However, these results are preliminary, and further validation on larger and more diverse datasets is required.

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

Abbreviation	Definition
ADC	Apparent Diffusion Coefficient
DW	Diffusion-Weighted
MRI	Magnetic Resonance Imaging
DKI	Diffusion Kurtosis Imaging
VERDICT	Vascular, Extracellular, and Restricted Diffusion for Cytometry in Tumors
IMPULSED	Imaging Microstructural Parameters Using Limited Spectrally Edited Diffusion
EES	Extra-cellular-extravascular
ic	Intra-cellular
ec	Extra-cellular
AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
ME	Multi-Exponential
LR	Linear Regression
PWLR	Piecewise Linear Regression
KNN	K-Nearest Neighbour
RFR	Random Forest Regressor
PW	Patch-wise
MLP	Multi-Layer Perceptron
PGSE	Pulsed-Gradient Spin Echo
RF	Radiofrequency
TE	Echo time
PDX	Patient-Derived Xenograft
PBS	Phosphate Buffer Saline
HE	Haematoxylin and Eosin
TR	Repetition time
WT	Wild type
NaN	Not a number
SNR	Signal-to-noise ratio
LOO-CV	LeaveOneOut-Cross Validation
MAE	Mean Absolute Error
CI	Confidence Interval

Abbreviation	Definition
MAPE	Mean Absolute Percentage Error
REL MAE	Relative Mean Absolute Error
NN	Neural Network

Symbols

Symbol	Definition	Unit
ADC	Apparent Diffusion Coefficient	[$\mu m^2/ms$]
K	Kurtosis	[-]
f_{ic}	Intracellular fraction	[-]
f_v	Vascular fraction	[-]
T_2	T2-relaxation time	[ms]
L	Cell size	[μm]
G	Gradient magnitude	[mT/m]
b	b-value	[s/mm^2]
A	Cell area	[μm^2]
l	Cell-wise diameter	[μm]
D_0	Intrinsic cytosol, i.e. intracellular, diffusivity	[$\mu m^2/ms$]
D_ex	Extracellular diffusivity	[$\mu m^2/ms$]
δ	Gradient duration	[ms]
Δ	Gradient separation	[ms]
t_{diff}	Diffusion time	[ms]
γ	Proton gyromagnetic ratio	[$s^{-1}T^{-1}$]

1

Introduction

Magnetic Resonance Imaging (MRI) is an important modality for imaging in oncology and is applied for various purposes. Several sources of contrast can be exploited to show characteristics of the imaged tissue, provided an appropriate MR-acquisition scheme is deployed. Diffusion-Weighted Magnetic Resonance Imaging (DW-MRI) is a quantitative MRI method, where the acquisition scheme is designed in such a way that the diffusion of water molecules is exploited as a source of contrast among tissues. By sensitising these DW-MRI measurements to water diffusion, we enable the noninvasive quantification of tissue microstructure [1–3]. Water molecules exhibit *Brownian motion* [4], i.e. a thermal molecular motion that causes fluctuations of their positions over time.

In free diffusion, particles move on average in an isotropic manner. Molecular displacement follows in this case a Gaussian distribution. In biological tissues however, the water molecules do not diffuse freely. As a result, the diffusion departs from being Gaussian distributed, often happening in a complex fashion. This deviation is due to the fact that the random walks of water molecules are restricted and/or hindered by various types of biological structures, such as cell membranes, collectively known as *tissue microstructure*. The ultimate goal of DW-MRI, and also of this thesis, is the estimation of the microstructure that has given rise to the observed DW-MR-signal [1, 5, 6].

The voxel size of *in vivo* MRI in humans is of the order of a few cube millimetres [7]. The observed MR-signal in a voxel arises from contributions of various diffusion processes taking place within different cellular components, since the voxel size is much larger than the scale of the microstructure where diffusion takes place ($\sim 1 - 100\mu m$). As a consequence, *in vivo* DW-MRI measurements are characterised by intrinsic, intra-voxel partial volume effects (i.e. *intra-voxel heterogeneity*), since the signal entangles the contribution of multiple water pools into one measurement. In Fig. 1.1 an illustration of intra-voxel heterogeneity is visualized.

The inference of microstructure from signal measurements is a challenging task [8, 9]. Different combinations of microstructural parameters can provide virtually indistinguishable signals in the presence of noise, making microstructure estimation an ill-posed inverse problem [10]. Different techniques have been proposed to disentangle such contributions, with the ultimate aim of obtaining sensitive and specific biomarkers of tissue microstructure [9, 11].

1.1. Research objectives

As the first part of this thesis, a literature review was executed to answer the following research question:

Research question 1: *Literature Review*

What is the current state of the art of DW-MRI methods for assessment of cancer microstructure in body imaging, and what is their value in the assessment of histology?

A number of recent articles has reviewed the state-of-the-art of DW-MRI applications [1, 12–16]. However, these tend to focus heavily on brain imaging, especially in neurological or psychiatric disorders, and reviews in contexts of body imaging in oncology are limited [7, 17–20], despite the rapidly growing amount of research in this field. The focus of this literature review was therefore on the state-of-the-art of DW-MRI, specifically in the context of oncological body imaging (beyond the central nervous system, i.e., mainly abdominal and pelvic imaging), aiming to analyse its value in the assessment of tissue microstructure as compared to reference co-localised histopathological measurements obtained on the same tissues.

This full literature review can be found in Appendix A, but will be summarised here. The articles included in the review show that measuring key biological properties, e.g., vascularisation, cell size, and cellularity (i.e. cell density), of tumours throughout the body in a clinical setting may be feasible with a variety of different tools.

From the literature review could be concluded that several of these DW-MRI techniques have the potential to enable the non-invasive estimation of distinctive microstructural properties of cancer in the body. Comparisons with histological features suggest that metrics from these methods provide histologically meaningful microstructure indices. These metrics, while not always accurate, may aid in equipping oncologists with non-invasive biomarkers of cancer aggressiveness. Despite the high potential, it is however observed that several challenges remain.

Most of the currently used methods for assessment of these microstructural properties can be categorized either as (i) *phenomenological signal representations*, e.g. the Apparent Diffusion Coeffi-

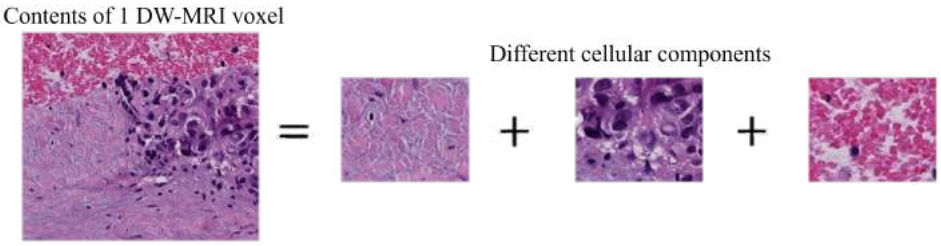


Figure 1.1: An illustration of *intra-voxel heterogeneity*: the voxel size of an *in vivo* DW-MRI in humans is much larger than the scale of the microstructure where diffusion takes place. Because of this, the signal observed in one DW-MRI voxel is the result of entangling the signal contributions of multiple different water pools into one single number.

cient (ADC) and Diffusion Kurtosis Imaging (DKI), or (ii) *multi-compartmental biophysical models*, e.g. Vascular, Extracellular, and Restricted Diffusion for Cytometry in Tumors (VERDICT) and Imaging Microstructural Parameters Using Limited Spectrally Edited Diffusion (IMPULSED).

The so-called signal representations are referred to as *phenomenological*, as they provide a description of the signal that in general does not enable the direct estimation of biophysical properties, such as cell size or cellularity. These approaches rather parametrise the signal on global, phenomenological properties of the whole displacement distribution of water molecules due to diffusion, e.g., the ADC or the apparent diffusion kurtosis K. The ADC is related to the mean squared displacement of water molecules. K on the other hand describes the level of non-Gaussianity of the overall displacement distribution, i.e., how different the diffusion profile is from ideal Gaussian diffusion as happening in free water. These properties are sensitive to alterations in the local microstructure due to pathology. However, they lack of specificity and may therefore be difficult to interpret, being surrogate indices that entail contributions from multiple biological factors in one number [21]. The ADC can for instance be thought of as the weighted average of the individual ADCs of different water pools, such as intra-cellular water, extra-cellular water, etc). Nonetheless, these measures are relatively easy to be obtained and the ADC has been widely utilised for multiple purposes, such as diagnostics, classification or grading of tumours, and therapy response assessment [22, 23].

Multi-compartment biophysical models describe the DW-signal as arising from the contribution of multiple water pools located in different cellular compartments. Fitting such models to sets of DW-measurements provides voxel-wise estimates of salient biophysical properties, such as the intra-cellular fraction (f_{ic}) or characteristic restriction size, effectively reflecting the mean intra-voxel cell size L. However, the use of models like these requires many assumptions on for instance the geometry of the tissue (e.g. modelling cells using various compartment shapes [24–26]), the number of tissue compartments (e.g. the vascular (v), intra-cellular (ic), extra-cellular-extra-vascular (EES) space [27], or the water exchange between them [28]. The number and characteristics of the tissue parameters that can be estimated depend on the assumptions made when building the biophysical model. The performance of the models in assessment of tumour microstructure strongly depends on these assumptions and the method used for fitting the model to the DW-MR-data.

In general terms, multi-compartment models typically describe the total MRI signal S of a voxel in the form of a combination of signals from elementary components, e.g.:

$$S = \sum_n S_n(microstructure; MRI\ sequence). \tag{1.1}$$

Above, S_n is the characteristic signal of the n-th component of the signal, as for example signal arising from water trapped within cells (intra-cellular signal), from flowing blood (vascular signal) or from water in the extra-cellular, extra-vascular space. S_n depends on the characteristic microstructural features of the component (e.g., the total amount of water relative to the component in question and the cell size for intra-cellular water). These microstructural features are in that case the unknowns of the mathematical problem and have to be estimated. This estimation is achieved by measuring the MR-signal multiple times, for different MRI acquisition settings. The model in Eq. (1.1) above is then fitted to the dataset made of the varying MR-settings and the corresponding varying MR-signal measurements.

Another method of estimating microstructural parameters that recently has been gaining momentum is to avoid relying on any functional form of the DW-signal (whether phenomenological or based on a biophysical model), but rather to use Artificial Intelligence (AI) to directly predict microstructure from raw DW-MR-signal data. In the literature, several applications of AI are reported for similar purposes,

showing promising results [29–31]. In this thesis we follow this innovative approach, to test whether it is possible to outperform state-of-the-art multi-compartmental, biophysical modelling of the DW signal, by using AI-algorithms that enable the estimation of microstructural parameters. In particular, using AI on paired examples of MRI signals and microstructural, histological measurements avoids the need of relying on analytical expressions of the MRI signal, whose formulation always comes with hypothesis on the link between signal and microstructure. The main research objective of this thesis is therefore defined as:

Main research objective

The development of a novel AI-based method to estimate microstructural parameters in the liver, specifically cell size L and intracellular fraction f_{ic}, given paired examples of DW-MRI signals and microstructural features as seen on histology, outperforming the estimation of microstructure by a well-established biophysical model.

Specifically, we want to develop a method based on pre-clinical data, to eventually be deployed on *in vivo* patient data, which is tested in the final part of this thesis.

To achieve this, a pre-clinical study is firstly performed. Various Machine Learning (ML) and Deep Learning (DL) algorithms are implemented, trained with DW-MR-data co-registered to histology, aiming to accurately predict the cell size L and intracellular fraction f_{ic}. A selection of different algorithms is used, ranging from very simple to more complex, to get a comprehensive overview of what kind of algorithm is most suitable for this specific purpose. We have DW-MR-data co-registered to histology-derived maps of cell size L and intracellular fraction f_{ic}, i.e. the ground truth, available of seven mouse livers. The algorithms can therefore be trained in a supervised manner for this pre-clinical setting. The goal of this part is to answer the following research question:

Research question 2: *Pre-clinical study*

Which of the implemented algorithms is most suitable to predict cell size L and intracellular fraction f_{ic}, in a pre-clinical setting?

As stated, we will attempt to predict the histological maps using the (i) raw DW-MR-data as input. In addition to this, we will also compute the (ii) ADC and K signal representations for this raw data, and use (iii) parametric maps computed by fitting a multi-exponential (ME) model (similar to VERDICT [27]) of cell size L_{ME} and intracellular fraction $f_{ic,ME}$. These three different input feature types lead to the third research question:

Research question 3: *Pre-clinical study*

What type of input feature is most suitable for prediction of cell size L and intracellular fraction f_{ic}, in a pre-clinical setting: (i) using raw DW-MRI-data, (ii) using ADC and K signal representations, or (iii) using parametric maps of cell size and intracellular fraction as computed by a ME-model?

Thirdly, the most promising AI-method will be deployed on clinical DW-MRI data. For this, we use the algorithm pre-trained on the pre-clinical data, as we do not have the histological slides available for patients (i.e., no ground truth). Considering this, it is not possible to compare the predictions voxel-by-voxel. However, there are biopsies available, and thus a global comparison of the predictions can be executed. As later will be explained further, it is only possible to use input type (iii) the parametric

maps of cell size L_{ME} and intracellular fraction $f_{ic,ME}$ for this purpose. In other words, we will use the pre-trained algorithm to 'refine' these parametric maps computed by the ME-model. This leads to the final research question:

Research question 4: *Clinical study*

What is the performance of the most promising pre-trained algorithm from the pre-clinical study, when testing on clinical patient data, to predict histological cell size L and intracellular fraction f_{ic}, using parametric maps of cell size and intracellular fraction as computed by a ME-model?

We will specifically try to compare the performance of the algorithm with the estimation by the ME-model itself, i.e. try to see if we can improve this estimation using the pre-trained algorithm.

1.2. Outline

Research question 1 is answered thoroughly in the literature review in Appendix A and is summarized in the previous section. In Chapter 2, background information is provided on the relevance of microstructural assessment in oncology, the fundamentals of DW-MRI, and the signal representations that we will use as input features in the pre-clinical study. The rest of the thesis includes three experimental approaches, of which an overview is visualized in Fig. 1.2. All experiments are executed using Python 3.9. The introduction, methods, results, discussion and a short conclusion of each experiment are described in separate chapters. The first experiment, a simulation study, is reported in Chapter 3. This simulation study is performed to ensure the initial implementation of the algorithms is correct. In the next experiment, in Chapter 4, the real pre-clinical data is utilised to answer *Research questions 2* and *3*. In the third experiment, described in Chapter 5, clinical data is used to test the performance of the pre-trained algorithms on real patients, answering *Research question 4*. Finally, in Chapter 6, the main research objective is evaluated and a general discussion and conclusion is provided.

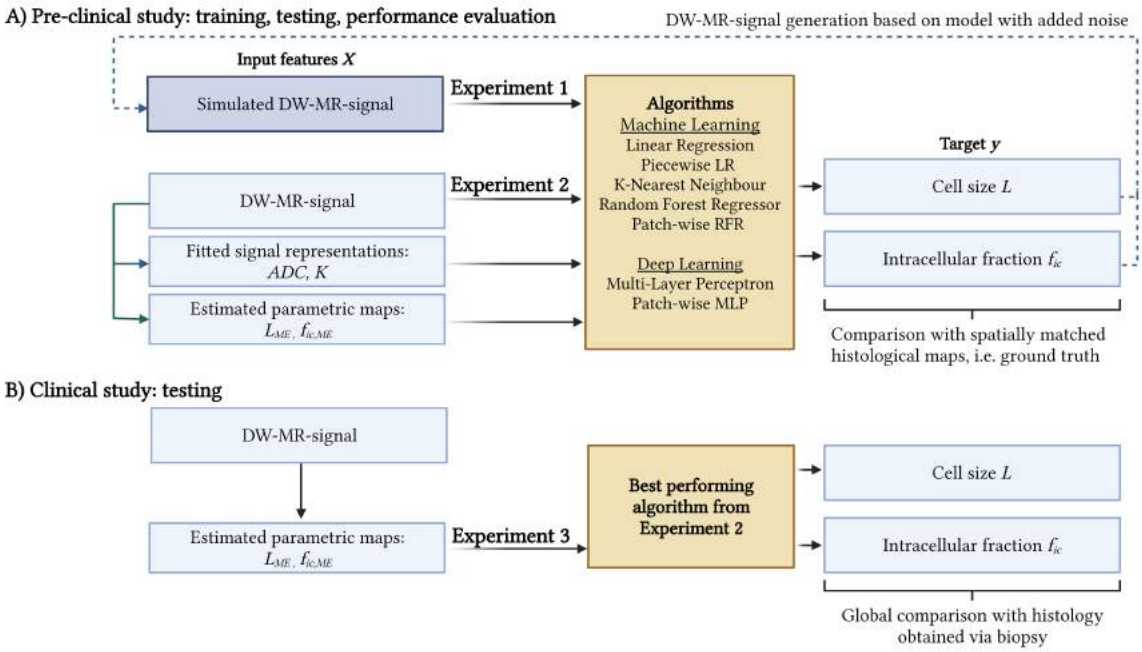


Figure 1.2: An overview of the methods used to achieve the main research objective, divided in three experiments. From top to bottom: (1) the simulation study, (2) the pre-clinical study, and (3) the clinical study. *Research questions 2* and *3* are answered in (2) the pre-clinical study. *Research question 4* is answered in (3) the clinical study. DW-MR = Diffusion-Weighted Magnetic Resonance, SNR = Signal-to-Noise Ratio, ML = Machine Learning, DL = Deep Learning, ADC = Apparent Diffusion Coefficient, K = Kurtosis, ME = Multi-Exponential.

2

Background

2.1. Microstructural assessment in oncology

Understanding the microstructure of a tumour can provide valuable insights about accurate diagnosis and treatment planning in oncology. One important characteristic of tumour microstructure is the intracellular fraction f_{ic}, which measures the proportion of the total volume of a voxel occupied by cells. This f_{ic} is closely related to tumour cellularity, which is a key feature of tumour aggressiveness. High cellularity is associated with faster tumour growth and a higher risk of metastasis, while low cellularity can indicate the presence of necrosis or fibrosis, influencing the efficacy of treatments [32]. Another important microstructural characteristic of tumours is the cell size, which can provide additional information about tumour aggressiveness and response to therapy. Heterogeneity of the cell sizes in a tumour often indicates neoplastic growth [33]. Specifically for liver cancers, acquiring information about tumour heterogeneity is found to be crucial, as it can reflect the prognosis and is predictive of treatment outcomes [34]. This information can be used to guide clinical decisions and treatment choices. Changes in cell size have been linked to treatment response, for instance in radiation therapy, which relies on the targeting of specific cell populations within the tumour [35]. It can also be used for the monitoring of treatment response, considering a significant change in cell diameter when cell death occurs [27].

2.1.1. Biopsy

Currently, the only way tumour microstructure and these histological properties can be quantified without excision of the tumour is via biopsy. In a biopsy, cells or tissues are removed from the body for examination by a pathologist. The use of biopsies however has several limitations. It is an invasive procedure for which contraindications exist [36] and might induce complications such as bleeding [37]. Additionally, during the biopsy procedure, there is a risk of seeding tumour cells into lymph nodes and/or into the vasculature, or even dragging tumour cells along the track of the biopsy needle. Both lead to an increased chance of the cancer spreading [38]. Furthermore, only a small part of the tissue can be sampled. This induces the risk of inadequately targeted biopsies, missing the tissue aiming to be sampled [39]. Finally, it is impossible to obtain an image of the microstructure of the tumour as a whole, making it difficult to capture the tumour's heterogeneity.

DW-MRI could function as a replacement of biopsy, to non-invasively quantify any tumour's mi-

crostructure as a whole and at any time, without risk of complications. This would be game-changing, not only for the diagnostics of cancer, but also for treatment response monitoring, where repeated imaging is required.

2.1.2. Histological validation of DW-MRI methods

As described, many approaches are proposed to estimate histological properties from DW-MR-data (Appendix A). Histological validation is an essential part, which is often not present in currently published studies. It is imperative to demonstrate that any new diffusion MRI metric is sensitive and biologically-specific to the microstructural property it intends to measure, before it can be used in the clinic. To this end, the voxel-by-voxel comparison of spatially-matched MRI and histology maps is the ideal way of performing exhaustive histological validation studies. Obtaining data of this type *in vivo* is however challenging, since it would imply performing histology of entire slabs of excised tissue, for example following surgical removal of tumours or the entire organ [40, 41]. Alternatively, tissue can also be imaged *ex vivo*, following excision. While on the one hand this would make MRI-histology alignment easier, on the other hand the tissue would lack of any perfusion, and microstructural changes would have occurred as a consequence of fixation, such as cell shrinkage. Lastly, comparing tiny 2D sections of histological material to 3D MRI has its own intrinsic limitations, since portions of tissues that contributed to the measured MRI signals are not included in the histological maps.

2.2. Diffusion-Weighted Magnetic Resonance Imaging

In DW-MRI, the MRI measurements are sensitised to water diffusion in the imaged tissue [1–3]. In the following sections, we describe the working principle and fundamentals of DW-MRI.

2.2.1. Diffusion encoding

Time-varying magnetic field gradients are used to sensitise the MRI signal to diffusion. The classical DW-MRI experiment is based on the Pulsed-Gradient Spin-Echo (PGSE) approach, also known as the Stejskal-Tanner experiment [42]. The schematic of a PGSE sequence is shown in Fig. 2.1.

The PGSE approach is based on the spin echo experiment, in which two radiofrequency (RF) pulses, of flip angle 90° and 180°, separated by a time $TE/2$ are used to irradiate the imaged tissue. The first pulse excites the tissue, creating a component of magnetisation orthogonal to the static field. The second pulse refocuses the magnetisation by cancelling out the effect of field inhomogeneities. The signal is sampled at a time TE, when the refocussing is complete, and the sampled signal is weighted by the underlying T2-relaxation time. In PGSE, two magnetic field gradient lobes (known as diffusion-encoding gradients) are added on either side of the refocussing pulse. The first gradient lobe effectively tags water molecule's phases depending on their spatial position, so that their phases will now depend on their position along the gradient direction. Conversely, the effect of the combination of the refocussing pulse and of the second gradient lobe is that the phase distribution modulation is cancelled out by the previous lobe, for those molecules that do not move. If all spins were perfectly static during diffusion-encoding, one would measure the same T2-weighted signal that one would have obtained without diffusion encoding. However, due to diffusion, water molecules change their position in between the two gradient lobes, as well as during the application of each lobe itself. This implies that a full phase coherence over the spin ensemble will not be re-established at the echo time $t = TE$, so that the signal measured will be smaller than the one obtained without diffusion-weighting, leading to signal attenuation. The amount of attenuation increases as the amount of diffusion taking place increases,

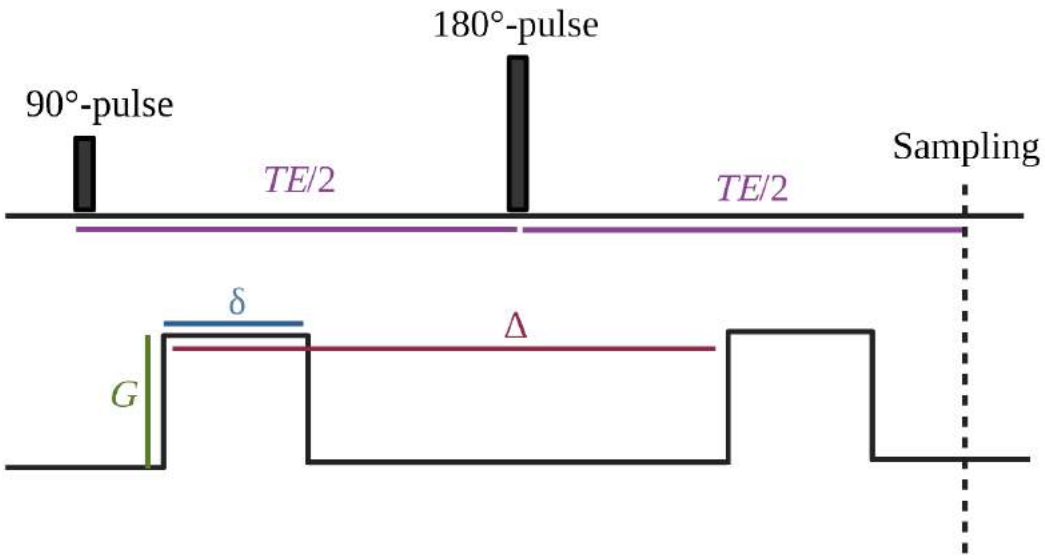


Figure 2.1: Schematic of the Pulsed-Gradient Spin-Echo (PGSE) sequence. The first and second radiofrequency (RF) pulses rotate the magnetisation vector by 90° and 180° respectively. These two pulses are characteristic for a spin-echo sequence, as they consecutively excite and refocus the magnetisation. After a time TE (the echo time), the signal is fully refocussed and sampled. The diffusion encoding gradient exists of two pulses on each side of the 180°. It is characterised by a gradient magnitude G, the gradient duration δ, and the separation time in between the two gradient lobes Δ.

resulting in DW images of lower intensity. An example of this is visualized in Fig. 2.2.

2.2.2. b-value and diffusion time

The amount of signal loss caused by diffusion-weighting depends on several factors. Some depend on the tissue being imaged, such as the underlying intrinsic diffusion coefficient and the characteristics of the microstructure, e.g., type and features of the biological structures that restrict or hinder diffusion. These features include, for example, the cellularity and the cell size. Other factors depend instead on the acquisition, and are: the gradient magnitude G, the gradient duration δ, and the gradient separation

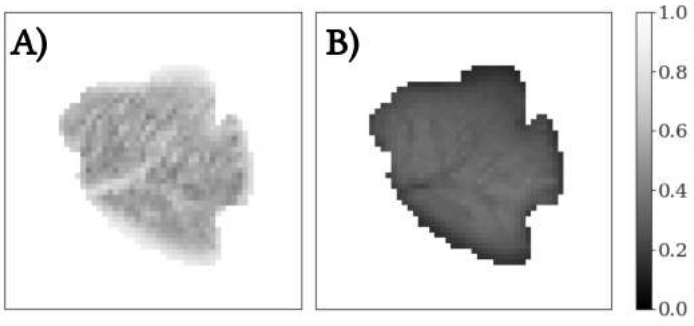


Figure 2.2: MR imaging of a healthy mouse liver *ex vivo* at room temperature. (A) Normalized MR-signal of a T2-image with zero diffusion weighting. (B) Normalized Diffusion-Weighted (DW)-image with $b = 918 s/mm^2$. The DW-image shows signal attenuation in areas of diffusion, resulting in a darker appearance in the image.

Δ (see Fig. 2.1). The total DW-signal S is the ensemble average of the signals from all spins [43], i.e.,

$$S \;=\; S_0 \;< e^{j\phi} >\tag{2.1}$$

where S_0 is the T2-weighted, non-DW-reference signal and ϕ is the phase accumulated by a spin. The ϕ depends on the interaction between a spin's random walk over time $\mathbf{r}(t)$ until the sampling time TE, and the temporal evolution of the diffusion encoding gradient $\mathbf{G}(t)$, i.e.,

$$\phi \;=\; -\gamma \int_0^{TE} \mathbf{G}(t) \cdot \mathbf{r}(t)\, dt,\tag{2.2}$$

where γ is the proton gyromagnetic ratio. For free water self-diffusion without any barriers, Eq. (2.1) simplifies to:

$$S \;=\; S_0 \;\mathrm{exp}(-b\, D_0).\tag{2.3}$$

Above, S_0 is the non-DW signal, whereas D_0 is the intrinsic diffusivity of water at the experiment temperature, and b provides a general indication of the overall strength of the diffusion-weighting. This factor depends on the acquisition settings, and is routinely known as b-value. It can be calculated as:

$$b = \gamma^2 G^2 \delta^2 (\Delta - \delta/3)\tag{2.4}$$

where γ is the proton gyromagnetic ratio, and G, δ and Δ the gradient parameters [44]. Another useful sequence parameter is the overall diffusion time $t_{diff} = \Delta - \delta/3$. This provides an indication of the amount of time that diffusing water molecules are allowed to experience the microstructure, before the MRI signal is acquired.

2.3. Signal representations

It can be shown that the signal in Eq. (2.1) can be expanded as a function of increasing powers of the b-value, i.e., $ln(S) \;\sim\; \sum_{k=1} a_k\, b^k$ (the cumulant expansion [45]), where coefficients a_k are related to the cumulants of the spin displacement distribution within a voxel. Expanding Eq. (2.1) up to the second power of b, provides (i.e., with an approximation error proportional to $O(b^3)$)

$$S = S_0\, e^{-b\, ADC \;+\frac{1}{6} K (b\, ADC)^2}.\tag{2.5}$$

In Eq. (2.5) above, ADC and K are respectively the apparent diffusion and excess kurtosis coefficients along the direction of the diffusion encoding gradient. ADC provides a measurement of the overall amount of diffusion taking place during the measurement, so that higher ADC implies stronger signal decay.

When the b-value is not too high (generally, not exceeding 400-800 s/mm^2 in body imaging or 1000-1500 s/mm^2 in the brain), the non-Gaussian diffusion characteristics (i.e., the excess kurtosis K) can be neglected, and the expansion reduces to

$$S = S_0 e^{-b\, ADC}\tag{2.6}$$

where S_0 is the non-DW-signal, b the b-value and ADC is the apparent diffusion coefficient. The difference between a mono-exponential (Eq. (2.6)) and a second order fit (Eq. (2.5)) including K is visualized in Fig. 2.3.

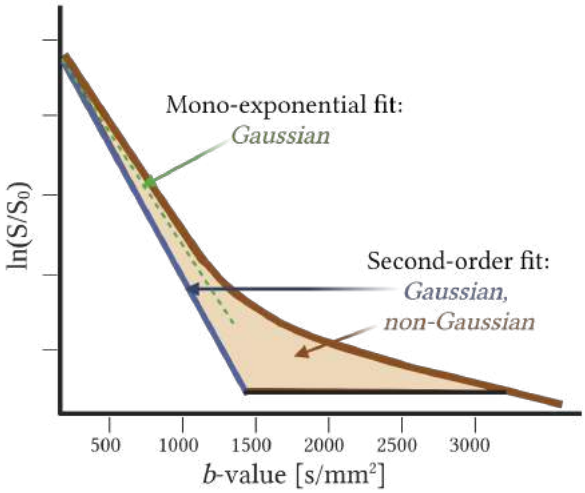


Figure 2.3: An example of the difference between a mono-exponential fit and a second-order fit including kurtosis. On the x-axis, the various b-values are plotted. On the y-axis, the logarithm of the signal attenuation due to the diffusion gradients is shown.

The ADC of Eq. (2.6) is easy to obtain, since it can be computed with as few as two images (one DW and one non-DW). It offers sensitivity to alterations in tissue microstructure due to pathology. For this reason, it is widely used for multiple purposes in the clinic, such as diagnostics, classification or grading of tumours, and therapy response assessment [22, 23]. However, as stated, it has a poor biological specificity, since contributions from different water compartments within the same voxel (e.g., intra-/extra-cellular water) are pooled together in one average number. This implies that the ADC measurement does not describe the complicated diffusion process in heterogeneous microstructures accurately. Signal contributions from different processes within a voxel, including pseudo-diffusion due to perfusion, are integrated and modelled by one single diffusion coefficient in Eq. (2.6), which also depends on the particular diffusion times used for the acquisition (e.g., gradient duration δ and gradient separation Δ). These factors make ADC only an 'apparent' coefficient.

The expanded version in Eq. (2.5) includes an extra variable K. By including this second-order term proportional to K, the equation accounts for non-Gaussian characteristics. In body imaging (e.g., prostate, bone, liver), it is common to evaluate scalar kurtosis indices, rather than full kurtosis tensors, due to relatively low diffusion anisotropy. Such kurtosis indices are typically reported in the form of an excess kurtosis K, so that $K = 0$ for Gaussian diffusion. While $K = 0$ would imply perfectly Gaussian diffusion, i.e., ideal mono-exponential signal decay, a non-zero K accounts for non-Gaussian diffusion. This can arise from a variety of intra-voxel sources, including the presence of multiple water pools with different intrinsic diffusivities, presence of restriction, water exchange, or combination of those [46, 47]. K is a unitless parameter. A larger K (in absolute value) implies greater departures from ideal Gaussian diffusion. Higher K values are typically observed in presence of higher cell densities, due to the increased level of diffusion restriction. [24]. The ADC in Eq. (2.5) is effectively a corrected estimate as compared to the first-order description $S = S_0 \, exp\,(-b\,ADC)$.

To measure kurtosis, it is necessary to sample the DW-signal at relatively high b-values (depending on the microstructure this is often higher than 1000 s/mm^2), at which departures from Gaussian diffusion become apparent.

3

Simulation study

3.1. Introduction

As described in Chapter 1, the relationship between the DW-MR-signal and the microstructure is complex, and the estimation of features such as cell size and intracellular fraction from MRI signal measurements is a challenging, ill-posed inverse problem. In this chapter, the implementation of algorithms developed to solve this inverse problem is described in detail. Moreover, results are presented regarding simulations conducted to verify the correct implementation of the algorithms and of the training routines.

To achieve this, simulated DW-MR-data is generated, of which we know that it has a clear relationship with the desired outputs. When feeding this data as input to the algorithms, it can be assumed that the algorithms should be able to predict the outputs with a high performance. The performance of the algorithms on this data can already be used as an indication of which algorithms are promising for the real data. Furthermore, we will analyse what hyperparameters appear suitable for this application.

To better understand the nature of the data that we will attempt to simulate, in the next section the real available pre-clinical dataset will firstly be described. The procedure for the data simulation is described in the section after that. Consequently, the procedure for training, testing, and evaluating the algorithms, and details of the implementation of the algorithms are described. Finally, the results and a short discussion are reported.

3.2. Methods

3.2.1. Pre-clinical dataset

The available data is anatomical (T2-weighted) and DW-MRI data co-registered with histological maps of seven mouse livers. Characteristics of the samples are described in Table 3.1. For three samples, WT, PDX_{inf1}, and PDX_{nec}, the T2W-image, the HE-staining and an example of the present microstructure is visualized in Fig. 3.1. Two of the same mouse livers (WT, PDX_{inf1}) were used for a study by Grussu et al. [25]. The mice were subcutaneously implanted with tumour tissue taken from patients, i.e. patient-derived xenografts (PDX), as illustrated in Fig. 3.2A. The tumour tissue gets a chance to develop further in the mice. This will occur for some mice more than others, allowing healthy liver to remain in some cases. After a time period, all mice are sacrificed and their livers resected. The

resected livers are used in both the *ex vivo* DW-MRI- and the histology procedure, as shown in Fig. 3.2B.

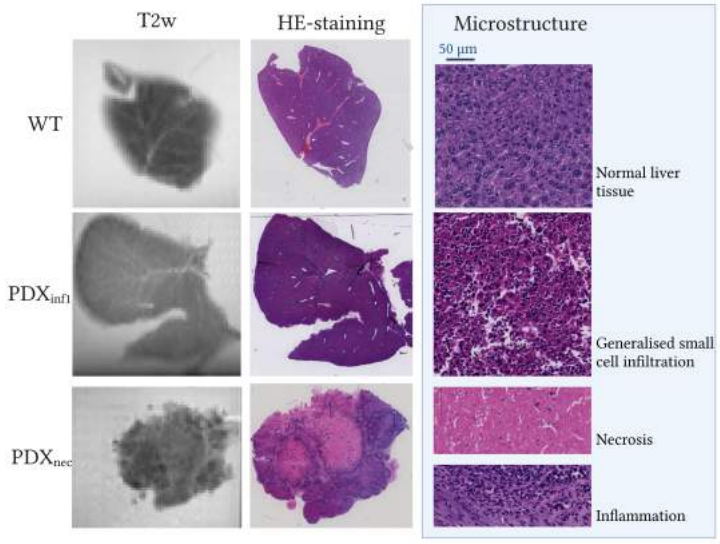


Figure 3.1: The pre-clinical dataset. Three representative samples are shown, from top to bottom: wild type (WT), i.e. healthy; Patient-Derived Xenograft (PDX) with generalized small cell infiltration, lymphoma-like (PDX_{infl}); PDX with necrosis and inflammation (PDX_{nec}). From left to right: a T2w-image, the histological section with HE-staining, and a zoom-in on the microstructure type most present in the sample.

DW-MRI

The (DW-)MRI-protocol was executed *ex vivo*, while immersed in a phosphate buffer saline (PBS) solution at room temperature. After MR-imaging, the livers were fixated and prepared for histological imaging, which is described in the next section.

The imaging acquisition protocol settings are as follows: a slice thickness of $2.2mm$ and resolution of $387\mu m \times 387\mu m$, based on PGSE, 9.4 T Bruker Scanner, $TR = 2700ms$, $\delta = 10ms$, $TE/\Delta = \{31/15,$ $45/15,$ $65/15,$ $45/30,$ $65/30\}$ ms, and 10 b-values spaced between $0 - 2800s/mm2$ for each TE/Δ combination: $b = [0.43, 1855.96, 119.17, 1062.44, 234.23, 917.94,$ $410.99, 1495.48, 721.63, 2792.19]s/mm^2$. This results in a total of 50 images for each subject. Two slices are imaged, resulting in a total size of $\{86, 86, 2\}$ voxels per sample.

Histology-derived parametric maps

The formation of the histological maps was already executed for a previous study by Grussu et al. [25]. After MRI, for both of the imaged slices, a $4\mu m$-thick histological section was attempted to be obtained. The slices were first stained with haematoxylin and eosin (HE), and consequently digitized with a Hamamatsu C9600-12 scanner (resolution: $0.277\mu m$). Using the QuPath [48] software for digital pathology image analysis, cells were segmented. Based on these segmentations, the cell area A is computed. Consequently, using:

$$l = \sqrt{\frac{4}{\pi}A}, \tag{3.1}$$

the cell-wise diameters l were determined. To obtain maps of cell size L and intra-cellular fraction f_{ic} on the same resolution as the MR-images, these diameters l were analysed by dividing the histological images into patches matching the MRI resolution of $387\mu m \times 387\mu m$. For each patch, the histological cell size L is computed by:

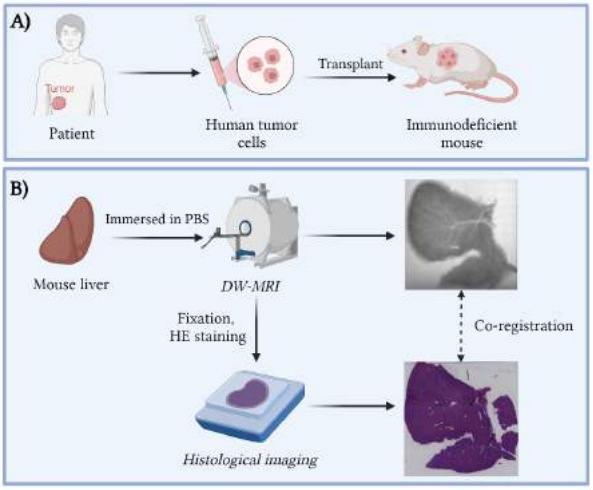


Figure 3.2: Acquisition of the pre-clinical dataset. (A) Patient-Derived Xenograft (PDX): mice are subcutaneously implanted with tumour tissue taken from liver cancer patients. (B) After a time period in which the tumour is allowed to develop further in the mice, they are sacrificed and their livers resected. The resected livers are immersed in phosphate buffer saline (PBS) solution and imaged with a DW-MRI-protocol *ex vivo*. After MR-imaging, the livers are fixated and prepared for histology. The slices were stained with haematoxylin and eosin (HE), and consequently digitized with a Hamamatsu C9600-12 scanner (resolution: $0.277\mu m$). Parametric maps of cell size L and intracellular fraction f_{ic} are derived from the histological images, and these are registered to the DW-MR-images.

$$L = \left(\frac{< l^7 >}{< l^3 >} \right)^{\frac{1}{4}}, \tag{3.2}$$

which is the so-called *volume-weighted cell size* [25]. The rationale for using this metric instead of the simple arithmetic mean $< l >$ is described in detail in [25, 49]. Briefly, the metric in Eq. (3.2) accounts for the fact that cells with a larger volume contain more water, and thus contribute more to the total MR-signal S_{tot}. S_{tot} is known to be the average of the cell signals weighted by cell volume, i.e., $S_{tot} = \frac{<l^3 S(b,l)>}{<l^3>}$, rather than $S_{tot} = < S(b,l) >$. Since $S(b,l) \sim e^{-\alpha l^4} \approx 1 - \alpha\, l^4$ [25], it can be shown that S_{tot} is roughly proportional to $\frac{<l^7>}{<l^3>}$. This implies that $\left(\frac{<l^7>}{<l^3>}\right)^{\frac{1}{4}}$ better corresponds to the cell size estimated through MRI than the simple arithmetic mean of the cell sizes within a voxel ($< l >$) [25, 49]. Therefore, the cell size L as computed by Eq. (3.2) is used as the ground truth that we want to predict. It should be noted that cells measured on histological sections are on average smaller than *in vivo*, because of two main reasons. The first is that cell shrinkage occurs during preparation for histology, due to dehydration, clearing, and paraffin embedding. The second reason is that we only see a 2D section from the 3D structure of the actual cells on the histological image. We rescale the L by $1.148\times$ and $1.273\times$, respectively to take into account the histological preparation and the bias from the 2D sectioning.

The intra-cellular fraction f_{ic} is computed by analysing the area within one patch that belongs within cells, and dividing this by the total patch area, i.e. the MRI voxel size. When f_{ic} is very low, i.e., there are little to no cells, it makes sense that the cell size in that voxel cannot be reliably estimated. To circumvent this, a threshold is set at $f_{ic} < 0.1$, so that for voxels with f_{ic} below this, the cell size L is set to Not a Number (NaN). In this way, these unreliable cell size estimates are not taken into account in training and testing.

By symmetric diffeomorphic registration of specimen's manual outlines [50], the histology-derived

Table 3.1: Description of the seven samples. Only voxels for which all types of data are available can be used in the experiments. Therefore, the images are masked and voxels for which one type of data is missing are removed. The number of voxels remaining for each sample is reported here. SD = Standard Deviation, WT = Wild type, PDX = Patient-Derived Xenograft, NA = Normal Appearing, nec = necrosis, L = cell size, f_{ic} = intracellular fraction.

Sample	Available voxels	Description	Mean L [SD] [μm]	Mean f_{ic} [SD] [-]
WT	911	Wild type, i.e. healthy	27.1 [1.61]	0.766 [0.157]
PDX_{NA1}	2883	PDX but with normal liver structures	26.8 [2.65]	0.828 [0.128]
PDX_{NA2}	1116	PDX but with normal liver structures	27.9 [2.78]	0.667 [0.172]
PDX_{inf1}	3594	PDX with massive small cell infiltration, lymphoma-like	20.8 [2.68]	0.833 [0.143]
PDX_{inf2}	2687	PDX with some cell infiltration, lymphoma-like, but to a lesser extent than PDX_{inf1}	27.6 [1.72]	0.797 [0.137]
PDX_{inf3}	1415	PDX with some cell infiltration, lymphoma-like, but to a lesser extent than PDX_{inf1}	29.4 [1.24]	0.780 [0.108]
PDX_{nec}	2240	PDX with extensive areas of necrosis and diffuse inflammation	25.9 [2.80]	0.543 [0.232]
$Total$	*14846*			

parametric maps were registered to the MRI. A description of the seven samples is provided in Table 3.1, including their mean cell size L and intracellular fraction f_{ic}.

Preprocessing the data

Masking In order to be able to use a voxel as a training sample, the DW-MRI data as well as the ground truth target data should be available for that voxel. Therefore, the data needs to be preprocessed to ensure this. A mask is created with ones on voxels where there are both DW-MRI signal intensity values and ground truth values based on the histological map for both cell size L and intracellular fraction f_{ic} available. The number of voxels that remains after this masking operation is the actual number of voxels that is available for the training and testing, shown in Table 3.1. The rest of the voxels, not belonging to the mask, will be set to NaN in the images. As mentioned, it was attempted to obtain histological images of both two DW-MR slices. However, this was not possible for two samples. For these (WT, PDX_{inf3}), there is only one histological slice imaged. Still, the availability of multiple slices for the rest of the samples will be utilised in the used cross-validation method, as described later in Section 3.2.3.

Normalisation Due to a lack of standardisation of DW-MRI acquisition and differences in machinery, it is required to remove the influence of the non-DW-signal S_0. To do this, the same method as in a study by de Almeida Martins et al. (2021) was utilised: we normalised the 50 DW-MR-acquisitions by dividing them by the signal value acquired at the lowest b-value and shortest TE-time [51]. Here, this means using the DW-MRI at $b = 0.43 s/mm^2$ and TE = 31 ms. Due to this normalisation, this first DW-MRI will become all ones in the image and will thus not contain useful information any more. However, the rest of the acquisitions will be normalized between 0-1. Some algorithms are not affected by normalisation, while others are. To keep the pipeline equal for all algorithms, however, normalisation of the data is applied for all algorithms.

Furthermore, sometimes the performance of algorithms also benefits from normalisation of the output. Intra-cellular fraction is, it being a fraction, already between 0-1, but the cell size ranges from $\sim 14{-}42 \mu m$. Therefore, the outputs are also normalised using a *MinMax*-scaler from the `scikit-learn` library, resulting in a normalisation between 0-1.

3.2.2. Simulating the data

The synthetic data is based on a deformation of the histological cell size L and intra-cellular fraction f_{ic}, using a multi-exponential (ME)-model. The model we use here ignores signal contribution of the vascular compartment. It assumes that the DW-MR-signal originates from two different compartments, intra-cellular and extra-cellular. This method is similar to the VERDICT model [27]. The total signal is described by:

$$S = f_{ic}exp[-bADC_{ic}(L, D_0, \delta, \Delta)] + (1 - f_{ic})exp[-bD_{ex}] \tag{3.3}$$

where f_{ic} is the intra-cellular fraction, D_{ex} the extra-cellular diffusion coefficient, D_0 the intrinsic cytosol diffusivity in $\mu m2/ms$, L the cell size in μm, Δ and δ respectively the gradient separation and duration in ms, and b the b-value in $\frac{ms}{\mu m2}$.

The intra-cellular ADC_{ic} depends on the latter four parameters, and is modelled by a Gaussian phase approximation for diffusion within a sphere [52]:

$$ADC_{ic} = \frac{2}{D_0\delta^2(\Delta - \delta/3)} \sum_{m=1}^{\infty} \frac{\alpha_m^{-4}}{\alpha_m^2 R^2 - 2} \left(\frac{(2\delta - 2 + e^{-\alpha_m^2 D_0(\Delta-\delta)} - 2e^{-\alpha_m^2 D_0\delta} - 2e^{-\alpha_m^2 D_0\Delta} + e^{-\alpha_m^2 D_0(\Delta+\delta)}))}{\alpha_m^2 D_0} \right) \tag{3.4}$$

where α_m is the m-th root of $J'_{3/2}(\alpha R) - \frac{1}{2}J_{\frac{3}{2}}(\alpha R) = 0$, and $J_{3/2}$ is the Bessel function of order $3/2$ and $J'_{3/2}$ is the derivative of the Bessel function of order $3/2$. This model thus describes a simplified relationship between L, f_{ic} and the measured DW-signal. No time-dependency is taken into account for the extra-cellular diffusion signal. To reduce complexity of the signal model, it is chosen to use a fixed TE-time ($TE = 45ms$), even though acquisitions with three different TE-times are available in the real data. The same Δ, δ, and b-values are used as in the real acquisition, as described in Section 3.2.1. This selection results in 20 different simulated acquisitions remaining.

Both intracellular diffusivity D_0 and extracellular diffusivity D_{ex} are fixed at set values, with D_0 and D_{ex} varying randomly between $1.8 - 2.5\frac{\mu m^2}{ms}$ and $1 - 2\frac{\mu m2}{ms}$, respectively. The D_0 and ground truth L computed from histology is given as input to Eq. (3.4), to obtain an estimate of ADC_{ic}. Consecutively, this ADC_{ic} is combined with the ground truth f_{ic} as computed by histology and D_{ex} in Eq. (3.3) to finally determine the S_{sim}. For all samples, synthetic DW-signals are generated based on this model. Finally, the same normalisation procedure as for the real data is executed (Section 3.2.1).

Noise The Rician distribution has been found to characterize the noise present in MRI [53]. MR-images are reconstructed from a real and an imaginary component based on the complex Fourier transform [54]. A complex MRI z is thus represented by:

$$z = z_{real} + jz_{imaginary} = (rcos(\theta) + n_1) + j(rsin(\theta) + n_2) \tag{3.5}$$

where z_{real} and $z_{imaginary}$ are the real and imaginary components, respectively. Both are independently corrupted by Gaussian white, i.e. zero-mean, noises n_1 and n_2. The r and θ describe the magnitude and the phase of the signal, respectively. To avoid phase artefacts, most commonly the magnitude images are used in MRI [53]. The magnitude of the MRI signal $|z|$ is defined as:

$$|z| = \sqrt{(rcos(\theta) + n_1)^2 + j(rsin(\theta) + n_2)^2}. \tag{3.6}$$

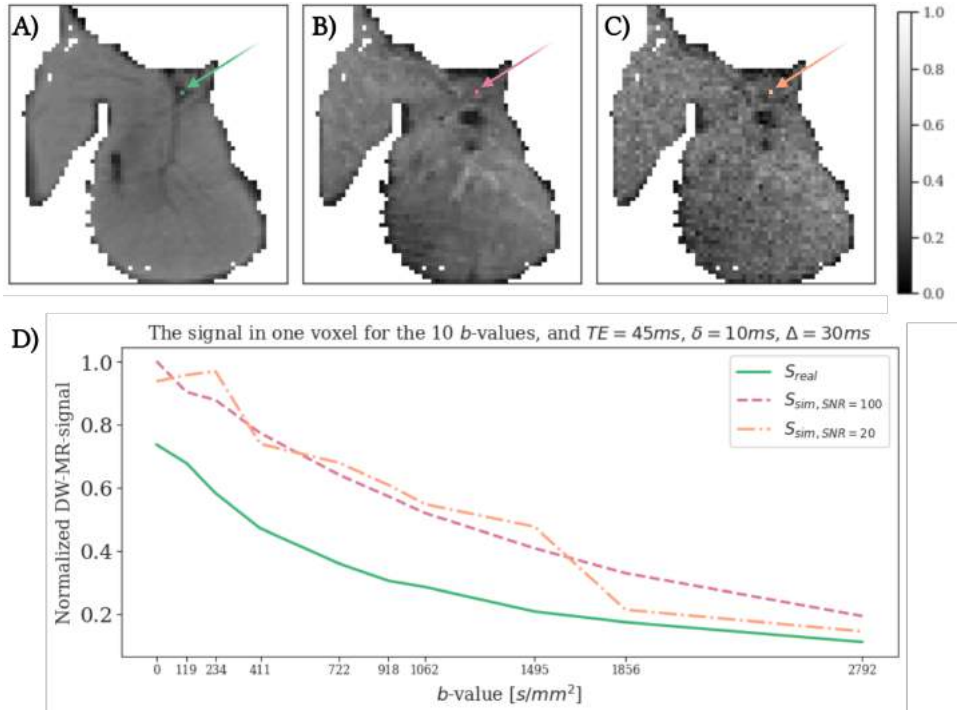


Figure 3.3: The normalized real and simulated data of one slice for sample PDX_{inf1} at $b = 1856 s/mm^2$, $TE = 45ms$, $\delta = 10ms$, $\Delta = 30ms$. A) The real DW-MR-image with S_{real}, B) the simulated image in ideal circumstances with $S_{sim,SNR=100}$, C) the simulated image in realistic circumstances with $S_{sim,SNR=20}$, D) the signal values for each b-value ($b = [0.43, 1855.96, 119.17, 1062.44, 234.23, 917.94, 410.99, 1495.48, 721.63, 2792.19]$ s/mm^2), $TE = 45ms$, $\delta = 10ms$, $\Delta = 30ms$. The voxel on which D) is based is indicated with the arrow in A-C.

.

Therefore, Rician noise is added to the synthetic signal according to this mathematical description of the signal. For simplification, it is assumed that the whole signal coming from the tissue is located in the real domain, resulting in:

$$S_{sim} = \sqrt{(S + (1/SNR) * n)^2 + (1/SNR) * n)^2} \tag{3.7}$$

where $rcos(\theta)$ is simplified to S: the simulated DW-MR-signal without noise, $rsin(\theta)$ is neglected, n is the noise, i.e. a random variable with a normal distribution centred at $\mu = 0$ and $\sigma = 1$ (Gaussian white noise), with the same dimensions as S. The n_1 and n_2 from Eq. (3.5) are assumed to be equal to n. Signal-to-noise ratios (SNR) of 100 and 20 will be used, to analyse the algorithms' performance in ideal SNR ($S_{sim,100}$) and realistic circumstances ($S_{sim,20}$), respectively. In Fig. 3.3, an example of the real and simulated data at the two SNR levels is visualized, for PDX_{inf1}, and the corresponding signal values for one voxel out of the shown slice.

3.2.3. Training, testing, evaluation

Each voxel is treated as an independent training sample, to ensure a set containing sufficient data points for training the algorithms. For each voxel, i.e. each sample i, a feature-target pair is available:

$$(\boldsymbol{X}^{(i)}, \boldsymbol{y}^{(i)}) \tag{3.8}$$

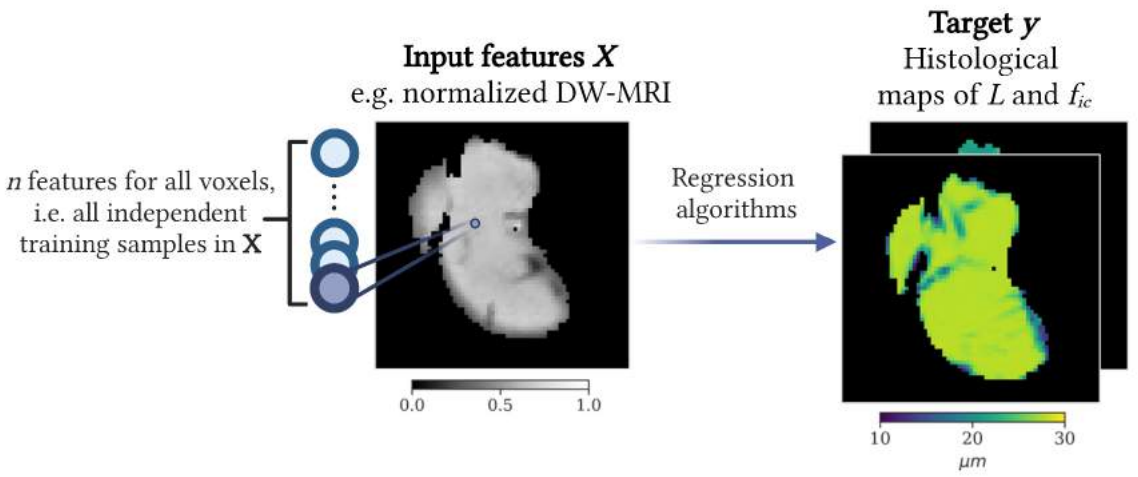


Figure 3.4: Illustration of the voxel-wise training approach. Each voxel is treated as an independent training sample. The voxels are used as input for seven regression algorithms, to predict the two separate outcomes. L = cell size, f_{ic} = intracellular fraction.

where $\boldsymbol{X}^{(i)}$ is the input feature vector containing n features, here including the 20 simulated DW-MR-acquisitions, and $\boldsymbol{y}^{(i)}$ the target vector, including the two outcomes we want to predict, cell size L and intra-cellular fraction f_{ic}. This voxel-wise training approach is illustrated in Fig. 3.4.

LOO-CV Leave-one-out cross-validation (LOO-CV) is used for the model training and testing, to ensure that the results are as generalizable as possible. In this way, the models are trained and tested in seven folds, with one subject in the test set for each fold. This is visualized in Fig. 3.5A.

As was visible from Table 3.1, the available data roughly contains three different types of tissue: normal liver structures (WT, PDX_{NA1}, PDX_{NA2}, PDX_{inf2}, PDX_{inf3}), lymphoma-like small cell infiltrations (especially in PDX_{inf1}, to a much lesser extent in PDX_{inf2} and PDX_{inf3}), and necrosis (PDX_{nec}). This implies that when either PDX_{inf1} or PDX_{nec} is located in the test set, the algorithms have no training data available that contains similar data for that specific microstructure type in the test set. In this way, the algorithms would never be able to learn the relationship between DW-MRI or the other input feature types and the resulting microstructure of the tissue, exhibiting small cell infiltration or necrosis. To circumvent this problem, we utilise the fact that we have multiple slices available for most subjects (5 out of 7 samples, and at least for the two unique samples, PDX_{inf1} and PDX_{nec}). After formation of the train and test sets for each fold, one slice is moved back from the test set to the train set (Fig. 3.5B). For two samples (WT, PDX_{inf3}) only one slice is available. For these samples, nothing changes in this adaption to the distribution: this slice remains in the test set.

Finally, for the one DL model, the Multi-Layer Perceptron (MLP), an additional split is made. The data remaining in the train set is split further into a validation set. This split is not done sample wise, instead, just the approximate equivalent of 1 of the slices is moved into the validation set; i.e. 7.15% of the voxels.

Evaluation For the slice remaining in the test set, the trained model is used to predict the outcomes, and these predictions are saved as *nifti* files for visual inspection of the performance. As described in the previous section, we essentially have three tissue types available, mainly occurring in WT,

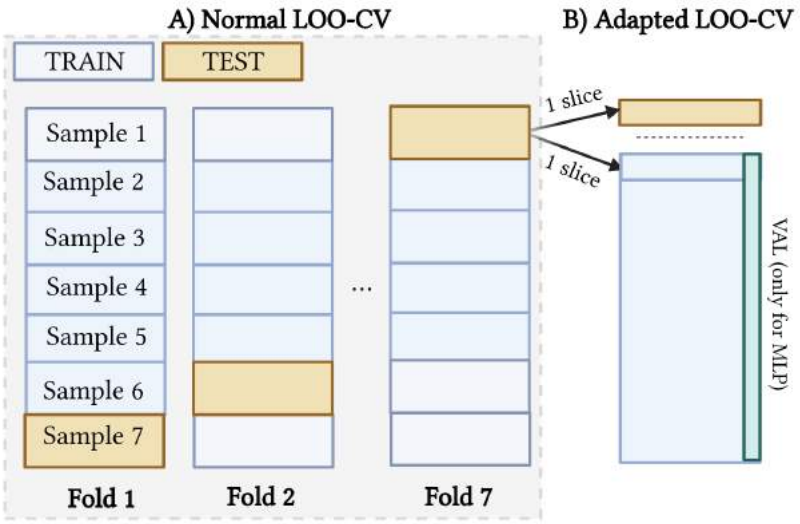


Figure 3.5: LeaveOneOut Cross-Validation (LOO-CV) is used in experiment 1 and 2, to ensure that the results are as generalizable as possible. (A) Because there are seven samples available, this means that there will be six training samples and one test sample in each fold. (B) As there is a limited amount of diversity in the data, we use an adapted version of the normal LOO-CV. After division into the train and test set, we move one of the slices in the test set back to the training set. This results in a test set with only one slice of the 'test' sample. For the Multi-Layer-Perceptron (MLP) we further split the training set into a train and validation set. This split is not done sample-wise, but on the shuffled whole set of voxels.

PDX_{inf1}, PDX_{nec} (Fig. 3.1). For this reason, these three samples are considered the *representative* samples, and these will be used for visual analysis of the performance.

For the seven predictions, two evaluation metrics are computed. The mean and $95\%CI$ are computed over the evaluation metrics of each fold.

The Mean Absolute Error (MAE) is a popular metric for regression problems [55]. For quantitative evaluation of the performance, the MAE will be computed with respect to the histology-derived ground truth. MAE is a simple metric and is computed as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{3.9}$$

where n is the total number of samples in the dataset, y_i and $\hat{y}_i$ are the true value and the predicted value respectively for the ith sample. The absolute value $|\cdot|$ ensures that the difference is always positive.

Mean Absolute Percentage Error (MAPE) is another popular performance metric for regression problem. This metric ranges from 0 to infinity and is defined by:

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}|\frac{(y_i - \hat{y}_i)}{y}|. \tag{3.10}$$

The advantage of this metric is that it does not depend on the units of the data, providing us with a percentage error. This causes the metric to be easy to understand and to compare between different outcomes. However, the metric fails or becomes extremely high if some of the actual values are equal or close to zero. Because this often occurs in the intra-cellular fraction f_{ic}, it being a fraction, the metric would provide a biased metric. To instead still provide an indication of how large the error is relative

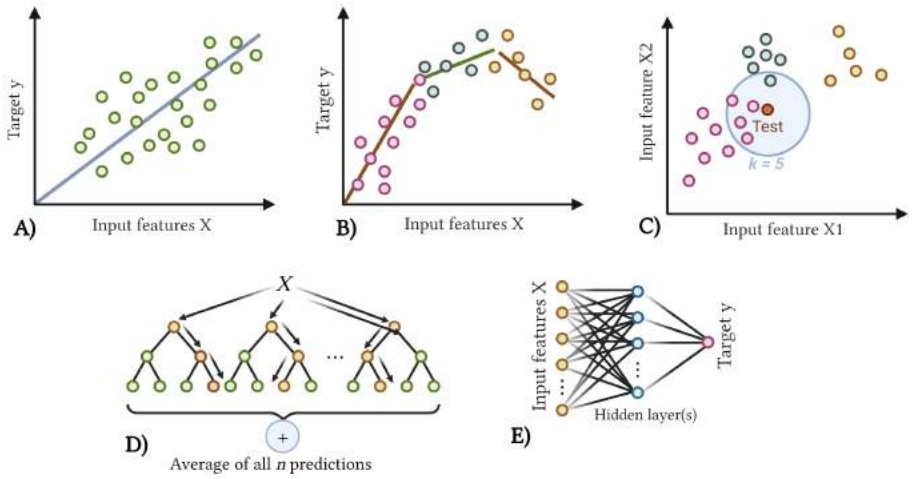


Figure 3.6: Illustrations of the implemented algorithms. (A) Linear Regression, (B) Piecewise Linear Regression, (C) K-Nearest Neighbour, (D) Random Forest Regression, (E) Multi-Layer Perceptron.

to the ground truth, a second MAE is computed [30]. This is done by normalising the output to [0,1], in the same way as described in Section 3.2.1. The *MinMax*-scaler is fitted to the ground truth values y and this scaler is used to transform both ground truth and the prediction. The resulting values are inputted in:

$$REL\ MAE = \frac{1}{n}\sum_{i=1}^{n} |y_{norm,i} - (\hat{y}_{norm,i})| * 100. \tag{3.11}$$

This provides a percentage instead of an absolute value. This metric will be utilised to compare the performance of the algorithms between the two outputs.

3.2.4. Algorithms

In this section, the algorithms implemented in the thesis are explained. The same algorithms will be used for the real pre-clinical study, described in the next study, in Chapter 4. The relationship between histology and DW-MRI-data depends on many factors and because of this, not a single model can be identified beforehand that will perform best. Various models are therefore tested, both ML- and DL-based, with an increasing complexity. The principle of the algorithms is illustrated in Fig. 3.6A-E.

Linear Regression

The most simple algorithm tested is Linear Regression (LR) (Fig. 3.6A). An implementation from the `scikit-learn` library is used, which is based on ordinary least squares LR. It finds the best fit by adjusting the coefficients $w_1, ..., w_p$ to minimize the difference between the predicted values and the actual values in the data.

Piecewise Linear Regression

Piecewise linear regression (PWLR) (Fig. 3.6B) is a type of regression analysis that involves fitting multiple linear models to different segments of the data, as opposed to a single linear model for the entire dataset like the LR in Fig. 3.6A. PWLR is useful when the data exhibits different linear trends

across different regions, as a single LR-model may not be able to accurately capture the underlying patterns in the data. Various methods can be applied for identifying how the data should be segmented. We utilise an implementation of `scikit-learn`, using a `DecisionTreeRegressor` for the segmentation. It is important to consider the number of segments that the feature space is eventually split in. When using a decision tree, this depends on various hyperparameters, like the depth of the tree. Various options are tested, and an optimal value is empirically chosen.

K-Nearest Neighbour Regression

K-Nearest Neighbours (KNN) regression works by finding the K nearest samples in the training data for each new (test) point (Fig. 3.6C). Consecutively, the target variable for this new point is predicted as the average of the target value of those K nearest points in the training data. The algorithm thus uses *feature similarity* to predict the values. Again, an implementation of the `scikit-learn` library is used.

The performance of the KNN regressor is mainly affected by the choice of the number of neighbours K (not to be confused with the kurtosis K) and the used distance metric. The chosen metric determines how the distance is computed between the new point and each training point. The three most popular distance metrics are *Euclidean, Manhattan* and *Minkowski*, where the latter is a generalized form of the first two. It is chosen to use the *Minkowski* metric. While a larger K can reduce noise in the predictions, it also leads to a smoother model. An optimal K is empirically identified.

Random Forest Regression

Random Forest Regression (RFR) is an ensemble learning method, i.e., multiple 'models' are combined, aiming to be more accurate than a single model. Specifically, RFR is based on an ensemble of *decision trees* (Fig. 3.6D). When using decision trees for regression, the input feature space is split into small regions. The mean of the target variable in that small region is then used as prediction for any new data point in that region. The algorithm works step-wise, where each time the feature and threshold are selected that result in the largest decrease in variance of the target variable. The space is then split along that feature and threshold, until certain stopping criteria are reached. The end result is a tree with a set of decision rules at the internal nodes and predictions at the leaf nodes. When a new data point is given to the model, a prediction is made by traversing the tree from root to a leaf node, using the decision rules at each internal node to determine the next node to visit.

Only using one decision tree as a regression model exhibits several problems, such as a high sensitivity to slight changes in data, which result in significantly different tree structures. Furthermore, decision trees struggle to find globally optimal solutions, and tend more to find a locally optimal solution. Using a whole forest of random decision trees, i.e. an RFR, the problems that single decision trees have, can be solved.

Within a random forest, trees do not interact, and they all predict a value independently. A predefined number of decision trees is constructed during training, and the mean is taken of the predictions of all the trees. For each new data point, each one of the decision trees will be used to predict the target value, and the average of all predictions is assigned to the final prediction. RFR has a high risk of overfitting, and because of that, hyperparameter tuning is important for this algorithm. By setting a maximal depth, you can control how deep the trees will grow. In addition, by setting the number of estimators, you can control how many trees are constructed, thus how large an ensemble you take at each step. An optimal value for both of these hyperparameters will empirically be determined.

Patch-wise RFR

In all the aforementioned algorithms, a voxel-wise approach is used. A disadvantage of this is that the spatial relationship between voxels is completely ignored. The surroundings of a voxel can however contain information about that specific voxel as well, and thus might be beneficial to integrate in the input given to the algorithm. Alexander et al. [56] researched this in 2017, for another application of parameter estimation from DW-MRI, and confirmed that the neighbourhood information is important and substantially improves the performance.

This is integrated in the patch-wise approach (Fig. 3.7), where patches are used to predict the central pixel of the patch. The image is divided in patches with a patch size n, using a *sliding window* over the voxels located within the mask. A step size of 1 is used to ensure that a patch is formed on every voxel. The output patch array is a square array of voxels including and surrounding the central pixel, with size $n \times n$. For voxels at or near the border of the mask, it is not possible to form a patch completely filled up with voxels containing information. These voxels are therefore not used in the training process, resulting in an erosion of $n - 1/2$ voxels at the border. Due to the still sufficiently large number of training voxels, this is not considered to be a problem. Then, for the remaining voxels, the mean of the values in the $n \times n$ patch is computed as a new feature. This results in two different feature types given to the algorithms as input: (i) the values for each feature (e.g. the 50 different DW-MR-acquisitions) in the central pixel, and (ii) the mean of the feature values over the $n \times n$ patch (also for e.g. the 50 different DW-MR-acquisitions). The final input feature vector for DW-MR-signal data in the patch-wise approach now has twice as many features as in the voxel-wise approach (100 vs. 50 features, respectively). In this way, the information of the surroundings is incorporated in the form of a mean value for the prediction of the central pixel.

When a patch size with $n > 3$ is used, an additional feature is computed in the form of (iii) the mean of the central 3×3 patch. So, for instance, when using a 5×5 patch, three features are used: the central voxel value, the mean over the whole 5×5 patch, and the mean over the central 3×3 patch. The two versions of the patch-wise approach are illustrated in Fig. 3.7.

This patch-wise approach is specifically used for the RFR, based on a similar approach by Alexander et al. (2017) [56]. The algorithm is implemented using the settings as described in the previous section, but now with the additional patch-wise feature(s).

Voxel-wise Multi-Layer Perceptron

A feedforward neural network is implemented using the `PyTorch` library. Specifically, a Multi-Layer Perceptron (MLP) is used, which is a traditional class of neural networks (NN) suitable for regression tasks [57]. An MLP is an NN with at least three layers: one input layer, one or more hidden layers, and one output layer, with any arbitrary activation function after each hidden layer. De Almeida Martins et al. [51] implemented two feedforward neural networks for parameter estimation based on DW-MRI, but for an application with many more input features (ranging $164 - 270$). In our study, a smaller number of input features is available, ranging from 2 for the parametric maps of the multi-exponential model to 100 for the patch-wise version of the DW-MRI. Because of this, a MLP is used with a similar network architecture as De Almeida Martin et al.'s shallower network, i.e. three fully connected hidden layers decreasing in number of nodes. However, in the first hidden layer a lower number of neurons will be used, corresponding to the lower number of input features. After each hidden layer, an activation function is used.

In addition to the network architecture, many other hyperparameters influence the performance of an MLP. The number of epochs is set to 30. Overfitting is prevented by the use of an early stopping

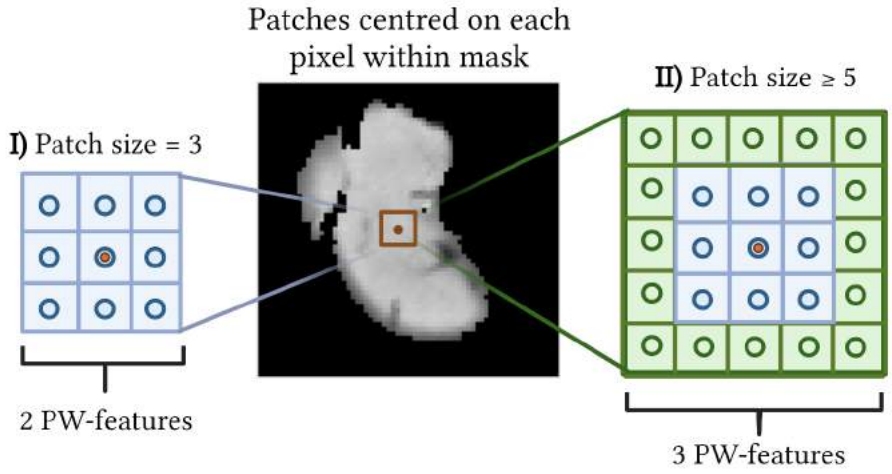


Figure 3.7: Illustration of the two versions of the patch-wise approach. If I) patch size $n = 3$, two patch-wise features are used: the intensity value at the central voxel and the mean over all voxels in the $n \times n$, while if II) patch size $n \geq 5$, three patch-wise features are used: again the intensity value at the central voxel, and the mean over all voxels in the $n \times n$ patch, and now with the mean over the voxels in the central 3×3 patch as additional feature.

method that terminated the training process when no decrease is observed in the validation loss for 3 (the *patience*) consecutive epochs.

The hyperparameters that are taken into account for optimisation in this case are: the learning rate, the batch size, and the network architecture.

Patch-wise Multi-Layer Perceptron

The MLP described in the above section is based on a voxel-wise approach. Because we are interested to see how the inclusion of neighbourhood information influences both ML- and DL-algorithms, the MLP is also implemented patch-wise. This simply implies the incorporation of the additional patch-wise feature(s) as input data for the MLP.

3.2.5. Implementation of the algorithms

As described in Section 3.2.4, there are hyperparameters to be optimised for almost all the algorithms. To do this, several hyperparameters are tested on the simulation, to see which values are appropriate for the application. For PWLR, the maximal depth was varied from $\{2, 4, 8\}$. The KNN was run with a number of neighbours of $K = \{10,\ 30,\ 80\}$. For the RFR, two parameters were analysed. The maximal depth of the trees was varied as $\{10,\ 20,\ 50,\ 100\}$. The number of estimators was varied as $\{8, 10,\ 20,\ 50\,100\}$. For the MLP, the learning rate is varied from $\{0.1, 0.01, 0.001, 0.0001\}$ and the batch size from $\{128, 256, 512, 1024, 2048\}$. Several model architectures are tested, with a varying number of hidden layers and neurons per layer. For the PW variant of the RFR and the MLP, the same hyperparameters were considered. Additionally, the performance using patch size 3×3 was compared to patch size 5×5 and patch size 7×7.

3.3. Results

3.3.1. Implementation of the algorithms

All models were implemented using the 20 simulated MRI-acquisitions, until it was clear that the model training functioned without bugs, by observance of an acceptable performance. The PWLR was implemented with a maximal depth of 4. The KNN-regressor was implemented using 30 neighbours. For the RFR, both voxel-wise and patch-wise, a maximal depth of 20 was chosen. Eight estimators (trees) were used. A batch size of 1024 and a learning rate of 0.01 were used. The final model architecture existed of three hidden layers with a decreasing number of neurons, $40 - 25 - 10$ respectively. After each hidden layer, a *ReLU*-activation function was used. For both the PW-RFR and the PW-MLP, a patch size of three was chosen.

3.3.2. Performance evaluation

The performance of the algorithms for prediction of cell size L and intracellular fraction f_{ic} is visualized in Table 3.2 and 3.3, respectively. Additionally, the test results for three representative samples are shown in Figure B.1 in Appendix B.

3.4. Discussion

The goal of this simulation study was to ensure the algorithms are implemented in a correct way, and to find hyperparameters suitable for this regression problem. To achieve this, we simulated data using a ME-model and added Rician noise.

Table 3.2: The results of the simulation experiment for prediction of the cell size L. MAE = Mean Absolute Error, REL MAE = Relative Mean Absolute Error, SNR = Signal-to-Noise Ratio.

	MAE [95% CI] [μm]		**REL MAE [95% CI] [%]**	
	$SNR = 100$	$SNR = 20$	$SNR = 100$	$SNR = 20$
LR	1.35 [0.652, 2.72]	1.37 [0.861, 1.82]	13.3 [8.35, 19.6]	14.7 [9.47, 22.7]
PWLR	1.03 [0.776, 1.36]	1.33 [1.03, 1.81]	8.78 [4.24, 12.3]	11.6 [4.99, 17.2]
KNN	1.07 [0.742, 1.67]	1.48 [1.04, 1.95]	9.1 [4.47, 13.9]	12.3 [7.39, 17.4]
RFR	0.88 [0.552, 1.38]	1.4 [1.12, 1.76]	9.51 [4.37, 13.2]	15.0 [10.6, 19.0]
PW-RFR	0.575 [0.313, 0.871]	0.936 [0.532, 1.55]	6.32 [2.46, 8.67]	10.1 [4.12, 14.0]
MLP	1.05 [0.453, 1.83]	1.21 [0.948, 1.9]	10.6 [5.56, 14.4]	12.9 [7.36, 16.4]
PW-MLP	0.542 [0.31, 0.82]	0.865 [0.48, 1.48]	5.92 [2.37, 8.39]	9.36 [3.78, 13.1]

Table 3.3: The results of the simulation experiment for prediction of the intracellular fraction f_{ic}. MAE = Mean Absolute Error, REL MAE = Relative Mean Absolute Error, SNR = Signal-to-Noise Ratio.

	MAE [95% CI] [-]		**REL MAE [95% CI] [%]**	
	$SNR = 100$	$SNR = 20$	$SNR = 100$	SNR = 20
LR	0.0764 [0.0549, 0.0995]	0.121 [0.0676, 0.158]	11.4 [7.97, 13.6]	18.1 [11.6, 23.6]
PWLR	0.061 [0.0452, 0.0947]	0.0934 [0.0489, 0.147]	7.42 [5.25, 11.2]	11.4 [5.67, 17.2]
KNN	0.0646 [0.0463, 0.122]	0.0967 [0.0607, 0.16]	7.81 [5.28, 14.2]	11.7 [7.02, 18.7]
RFR	0.0555 [0.0309, 0.0912]	0.0933 [0.0461, 0.139]	8.11 [5.58, 12.1]	13.7 [8.31, 18.4]
PW-RFR	0.041 [0.0216, 0.0663]	0.0658 [0.0293, 0.125]	6.05 [3.4, 8.81]	9.55 [4.61, 16.6]
MLP	0.0583 [0.0324, 0.0967]	0.0733 [0.0293, 0.165]	8.57 [5.75, 13.1]	10.6 [5.23, 22.1]
PW-MLP	0.0344 [0.0185, 0.0556]	0.0575 [0.027, 0.122]	5.03 [2.9, 7.39]	8.34 [4.23, 16.2]

Based on the results of this experiment, the patch-wise algorithms appear to be the most promising. A clear increase in performance is observed when the additional patch feature is incorporated for both the RFR (e.g. at $SNR = 20$, $REL\ MAE$ from 15.0 to 10.1% for L, and $REL\ MAE$ from 13.7 to 9.55% for f_{ic}) and the MLP (e.g. at $SNR = 20$, $REL\ MAE$ from 12.9 to 9.36% for L, and $REL\ MAE$ from 10.6 to 8.34% for f_{ic}).

The overall best performing is the most complex model, the PW-MLP, with the PW-RFR being a close second. However, using the simulated data, even the most simple model, LR, shows an acceptable performance (e.g. at $SNR = 20$, $REL\ MAE = 14.7$% and 18.1% for L and f_{ic}, respectively).

No large differences are yet observed between the prediction of cell size L vs. intracellular fraction f_{ic}, with for SNR = 20 the $REL\ MAE$ ranging from 5.92-13.3% for L and 5.03-11.4% for f_{ic}.

3.5. Conclusion

From the experiment described in this chapter, we conclude that the algorithms and their training routines are all implemented correctly. Each algorithm is able to predict the desired outputs with high performance. Furthermore, suitable hyperparameter settings were found using the simulated data. In the next chapter, the algorithms will be deployed on the real pre-clinical data, using the found hyperparameters.

4

Pre-clinical study

4.1. Introduction

In this chapter the pre-clinical experiment is presented, aiming to analyse which of the algorithms performs best in predicting microstructure based on DW-MR-data, and which input feature type is most suitable for this application.

The algorithms as described in the previous chapter will be trained again on the real data, using the best functioning hyperparameters as found in Experiment 1. The algorithms will be trained separately six times, i.e. using the three types of input features X (Table 4.1, Fig. 4.1) and the two targets y. The same voxel-wise approach as illustrated in Fig. 3.4 in the previous chapter is used for the training, including the LOO-CV adaptation (Fig. 3.5). For each algorithm the MAE and $REL\ MAE$ will be computed and the prediction on the slice remaining in the test set is saved for visual inspection. The results of this section will first be used to provide an answer to *Research question 2*:

Research question 2: *Pre-clinical study*

Which of the implemented algorithms is most suitable to predict cell size L and intracellular fraction f_{ic}, in a pre-clinical setting?

In order to objectively assess this, mainly the quantitative results are used, besides visual inspection. Consecutively, the best performing algorithm will be chosen, and analysed further to answer *Research question 3*:

Research question 3: *Pre-clinical study*

What type of input feature is most suitable for prediction of cell size L and intracellular fraction f_{ic}, in a pre-clinical setting: (i) using raw DW-MRI-data, (ii) using ADC and K signal representations, or (iii) using parametric maps of cell size and intracellular fraction as computed by a ME-model?

The use of the two additional input feature types *(ii, iii)*, handcrafted from the raw DW-MRI data, could have several potential benefits over the use of *(i)* the raw DW-MRI data. Firstly, each of the feature types provides unique information about the tumour biology. The raw DW-MRI data reflect the

Table 4.1: Description of the available input feature types. The three input feature types are used in the pre-clinical experiment. ADC = Apparent Diffusion Coefficient, TE = echo time, ME = Multi-Exponential.

Input feature type	Number of features	Description
(i) DW-MR-signal	50	Different acquisitions, including T2W- and DW-images with various b-values and TE-times.
(ii) $ADC - K$	10	Signal representations fit to the raw DW-MR-signal. The apparent diffusion coefficient and the excess kurtosis are fit for five different combinations between TE and diffusion time.
(iii) ME-estimations	2	Parametric maps of cell size and intracellular fraction as estimated by fitting a ME-model.

overall diffusion properties of the tumour, while ADC and K are often used in an attempt to provide more specific information about the cellularity and size of tumour cells. The ME-model allows us to incorporate biophysical modelling into the regression problem, which, as described in Chapter 1, has limitations, but also shows promising results. Secondly, Table 4.1 shows that using either one of the two handcrafted feature types induces a reduction in feature dimensionality: 50 to 10 to 2 features for the raw DW-MR-signal, $ADC - K$, and the ME-estimations, respectively. Using a lower number of input features allows for a simpler and more interpretable model that is less prone to overfitting. Thirdly, comparing the performance of the algorithms for different input feature types can help identify which type is the most informative for this prediction.

Additionally, using the parametric maps obtained from the ME-model can enable translation to clinical data, while the other input feature types may not be directly applicable in a clinical setting due to their dependence on the specific MR-acquisition protocol. This does induce the requirement that the parametric maps of cell size and intracellular fraction as computed by the biophysical ME-model should be independent of the input MR-protocol. This is especially relevant considering that the protocol utilised for our pre-clinical dataset is not reproducible in a clinical setting.

The synthesis of input feature types *(ii, iii)* is described in the following sections. Consecutively, the results, a short discussion and conclusion are reported.

4.2. Methods

4.2.1. Dataset

Three different types are analysed as input features X, to test which of them is most suitable. Firstly, the DW-MRI signal as described in Section 3.2.1 will be utilised as feature type *(i)*. Two other input feature types are formed based on this raw signal data, for which the computation will be described in the next paragraphs. The DW-MR-signal data is normalized using the method described in Section 3.2.1, by dividing all acquisitions over the lowest b-value and shortest TE-time image. Feature type *(ii)* and *(iii)* are normalised between 0-1, by using the mentioned *MinMax*-scaler. Cell size L and intracellular fraction f_{ic} are again used as targets y, based on the same acquisition process as described in Section 3.2.1. An overview of the available input features is shown in Table 4.1 and visualized in Fig. 4.1.

***(ii)* Signal representations** Two signal representations are fit to the DW-MR-signal, the ADC and the K, which were introduced in Chapter 1. As stated, the ADC describes the overall amount of diffusion in a voxel taking place during DW-MR measurement; a higher ADC implies stronger signal decay. ADC has been utilised for multiple purposes, such as diagnostics, classification or grading of tumours, and

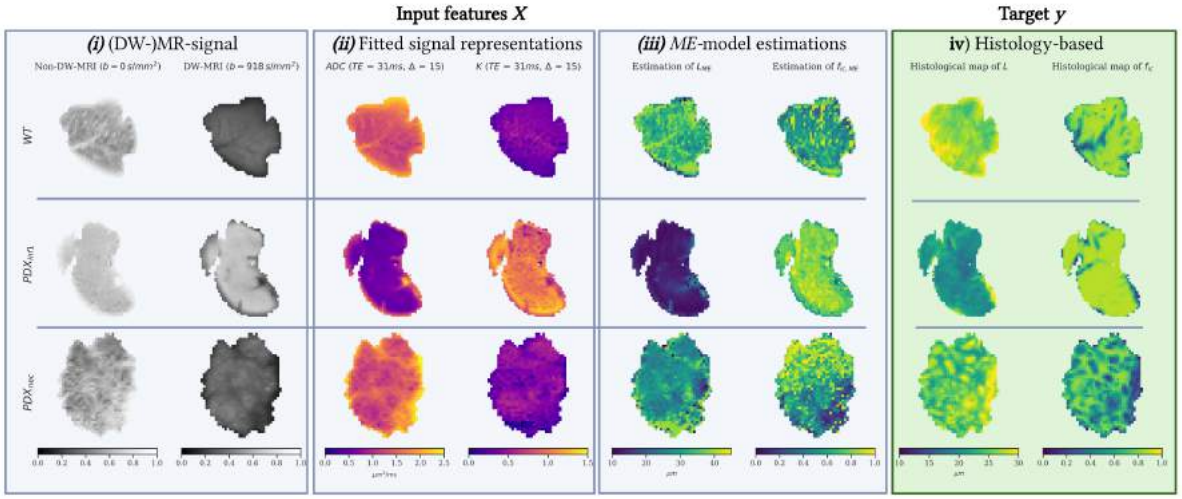


Figure 4.1: The available input (i-iii) and target data (iv) for three of the seven samples. From top to bottom: WT, wild type, i.e. healthy; PDX_{inf1}, PDX with massive small cell infiltration (lymphoma-like); PDX_{nec}, PDX with extensive areas of necrosis and diffuse inflammation. From left to right: (i) the normalized MRI, one non-DW, one DW; (ii) the maps of the fitted signal representations ADC and K; (iii) parametric maps of cell size and intracellular fraction estimated by a ME-model; (iv) the prediction target, i.e. the ground truth, parametric maps of the histology-derived cell size L and intracellular fraction f_{ic}.

therapy response assessment [58]. Diffusion Kurtosis Imaging (DKI) is a technique originally proposed by Jensen et al. (2005) [46]. The kurtosis K indicates how much the observed diffusion processes depart from Gaussian free diffusion.

ADC and K are both *phenomenological* signal representations, as they describe the MRI signal based on the statistical moments of the spin displacement distribution. They are surrogate indices that entail contributions from multiple biological factors from within one voxel into one single number [21]. Because of this, these metrics may be hard to interpret and do not provide direct estimates of microstructural characteristics. Even though they have a relatively poor biological specificity, they can still function as sensitive markers for tissue microstructure. The relationship between ADC, kurtosis K and microstructural characteristics has been researched extensively (e.g. in [22, 23, 25, 59–63] or see the Literature Review in Appendix A). Typically, ADC is negatively correlated with cell density, as a higher cell density leads to an increase in restriction [22]. K values are typically positively correlated with cell density, due to the increased level of diffusion restriction leading to greater departures from ideal Gaussian diffusion. In addition, K can carry information about the heterogeneity within a voxel [24].

These features were computed from the DW-MR-signal, by fitting a ADC-K model from the `MRItools` collection [64], based on Eq. (2.5), described in Chapter 2. It should be noted that the estimation of K is error-prone, considering that K becomes unstable for low ADC (approaching 0), as K is obtained by dividing the second cumulant by ADC and noise is thus amplified.

(*iii*) Parametric maps Furthermore, parametric maps of the cell size and intra-cellular fraction as computed by a multi-exponential ME-model are included as input feature type. This model is similar to the one used for the simulated DW-MR-signal generation in Section 3.2.2, but with several adaptions.

We derived maps of cell size and intra-cellular fraction following a two-step procedure. Firstly, we fitted a model that describes the total signal as the sum of vascular and non-vascular (tissue) compo-

nents, i.e., S:

$$S = S_0 \left(f_v e^{-bADC_V - \frac{TE}{T2_v}}) + (1 - f_v)(e^{-bADC_T + \frac{1}{6}b^2 ADC_T^2 K - \frac{TE}{T2_T}} \right) \tag{4.1}$$

where S_0 is the non-DW-signal, b the b-value, and TE the echo time. The first exponential describes the signal coming from the vascular compartment (V, this was not considered in the previous chapter), with f_v the vascular fraction, and the second exponential describes the tissue signal (T), i.e. everything non-vascular. The ADC_v and ADC_T respectively describe the apparent diffusion coefficient for the vascular and tissue compartment. The $T2_V$ and $T2_T$ describe the T2-relaxation times for the respective compartments, and K the kurtosis observed in the tissue.

Once the model in Eq. (4.1) was fitted, we performed a second fitting only including measurements where the vascular signal can be reasonably considered to have vanished. Here this means that the second fitting was performed only on acquisitions with $b > 1000 s/mm^2$. This is based on the fact that in our samples vessels are mainly filled with PBS, and there is no pulsating blood (diffusivity of the PBS being $ADC_V \simeq 2.4 \frac{\mu m^2}{ms}$). The second fitting tries to explain the tissue signal as the sum of the intra-cellular and extra-cellular, extra-vascular components, i.e.:

$$S = S_0(1 - f_v)e^{-T2_T} \left(f_{ic} e^{-bADC_{ic}} + (1 - f_{ic})e^{-bADC_{ec}} \right). \tag{4.2}$$

Here, ADC_{ic} and ADC_{ec} are the intra-cellular/extra-cellular ADC, and it is assumed that the T2-relaxation times are equal for both compartments, i.e. $T_{2,ic} \simeq T_{2,ec} \simeq T_{2,T}$. For the ADC_{ic}, Eq. (3.4) as defined in Section 3.2.2 is used. In contrast with the VERDICT model, the choice is made to fit all the remaining free variables to the data, and refrain from fixing any of them to a set value. Eq. (4.2) is fit to the DW-MR-signal data using a custom-written Python script for maximum-likelihood fitting, resulting in parametric maps with an estimation of cell size and intracellular fraction.

The use of the estimations of cell size and intracellular fraction by the ME-model is two-fold. We use them as input feature type *(iii)*, and in this case we will refer to the two of them as the ME-estimations. Furthermore, we want to compare the performance of our own implemented AI-algorithms to the estimations by this very biophysical ME-model. When we talk about the estimations of the model as this separate gold-standard model that we try to outperform, we will refer to them as L_{ME} and $f_{ic,ME}$. For performance comparison, the MAE and $REL\ MAE$ are computed for the estimation by the ME-model.

4.3. Results

The quantitative results for prediction of L and f_{ic} are reported in Tables 4.2 and 4.3, respectively. The performance metrics for the ME-model are visualized in these tables in the last row. The predictions for three representative samples for all seven algorithms and three input types can be found in Appendix C.

Table 4.2: The MAE and $REL\ MAE$ for prediction of cell size L, as predicted by the seven different algorithms, based on three input types. For comparison, the MAE and $REL\ MAE$ for the estimation by the ME-model is reported in the last row. MAE = Mean Absolute Error, REL MAE = Relative Mean Absolute Error.

	MAE [μm] [95% CI]			**MAE REL [95% CI] [%]**		
	DW-MR-signal	$ADC-K$	ME-estimations	DW-MR-signal	$ADC-K$	ME-estimations
LR	5.08 [1.37, 19.3]	2.95 [1.68, 6.5]	3.24 [1.51, 7.02]	32.3 [9.99, 90.3]	23.1 [11.2, 38.9]	25.8 [12.0, 52.4]
PWLR	2.8 [1.45, 5.29]	3.27 [1.49, 7.34]	2.89 [1.29, 6.77]	21.6 [13.7, 29.4]	25.3 [11.7, 41.0]	21.4 [12.1, 33.4]
KNN	3.12 [1.47, 8.01]	2.98 [1.77, 7.1]	2.89 [1.31, 6.54]	22.3 [13.8, 38.4]	22.5 [11.2, 35.1]	21.7 [12.5, 32.4]
RFR	3.11 [1.45, 8.22]	3.05 [1.64, 7.43]	2.88 [1.45, 6.68]	29.2 [17.1, 58.3]	29.2 [21.1, 52.2]	27.9 [17.8, 47.3]
PW-RFR	2.09 [1.45, 3.31]	2.19 [1.5, 2.72]	2.23 [1.3, 3.91]	23.4 [11.5, 43.0]	24.0 [12.1, 33.5]	24.8 [12.1, 50.5]
MLP	3.66 [1.14, 12.8]	3.38 [1.18, 10.5]	3.25 [1.1, 10.5]	32.4 [13.0, 89.1]	31.0 [13.5, 74.4]	29.3 [12.6, 73.1]
PW-MLP	2.07 [0.971, 3.12]	2.09 [1.33, 3.11]	2.23 [1.27, 3.7]	22.0 [11.4, 35.7]	22.2 [15.1, 35.5]	23.2 [15.2, 42.1]
L_{ME}		7.44 [6.38, 8.32]			8.61 [7.84, 9.36]	

Table 4.3: The MAE and $REL\ MAE$ for prediction of intracellular fraction f_{ic} as predicted by the seven different algorithms, based on three input feature types. For comparison, the MAE and $REL\ MAE$ for the estimation by the ME-model is reported in the last row. MAE = Mean Absolute Error, REL MAE = Relative Mean Absolute Error.

	MAE [-] [95% CI]			**MAE REL [95% CI] [%]**		
	DW-MR-signal	$ADC-K$	ME-estimations	DW-MR-signal	$ADC-K$	ME-estimations
LR	0.386 [0.105, 1.53]	0.138 [0.0766, 0.197]	0.15 [0.0888, 0.209]	45.4 [12.0, 175.0]	16.5 [9.47, 23.1]	18.0 [11.9, 24.4]
PWLR	0.159 [0.108, 0.22]	0.165 [0.0891, 0.29]	0.169 [0.113, 0.206]	19.6 [12.7, 31.1]	19.7 [11.2, 33.3]	20.3 [15.9, 24.1]
KNN	0.186 [0.0678, 0.564]	0.157 [0.0989, 0.206]	0.17 [0.116, 0.209]	21.9 [8.0, 64.5]	18.8 [13.1, 24.1]	20.4 [16.4, 24.4]
RFR	0.167 [0.0694, 0.395]	0.153 [0.103, 0.194]	0.166 [0.107, 0.218]	25.0 [12.3, 59.9]	23.0 [15.2, 28.6]	24.9 [16.9, 35.9]
PW-RFR	0.0994 [0.0523, 0.149]	0.117 [0.0733, 0.168]	0.127 [0.0597, 0.194]	14.8 [9.19, 23.5]	17.7 [11.0, 24.6]	19.4 [8.86, 31.6]
MLP	0.11 [0.0685, 0.195]	0.0861 [0.0496, 0.18]	0.0966 [0.0672, 0.18]	16.5 [10.2, 26.8]	12.7 [7.01, 24.2]	14.2 [9.92, 24.2]
PW-MLP	0.0878 [0.0372, 0.166]	0.0773 [0.0463, 0.181]	0.0756 [0.0272, 0.186]	12.8 [6.51, 22.2]	11.2 [6.6, 23.9]	10.8 [4.66, 24.6]
$f_{ic,ME}$		0.251 [0.11, 0.46]			0.302 [0.16, 0.51]	

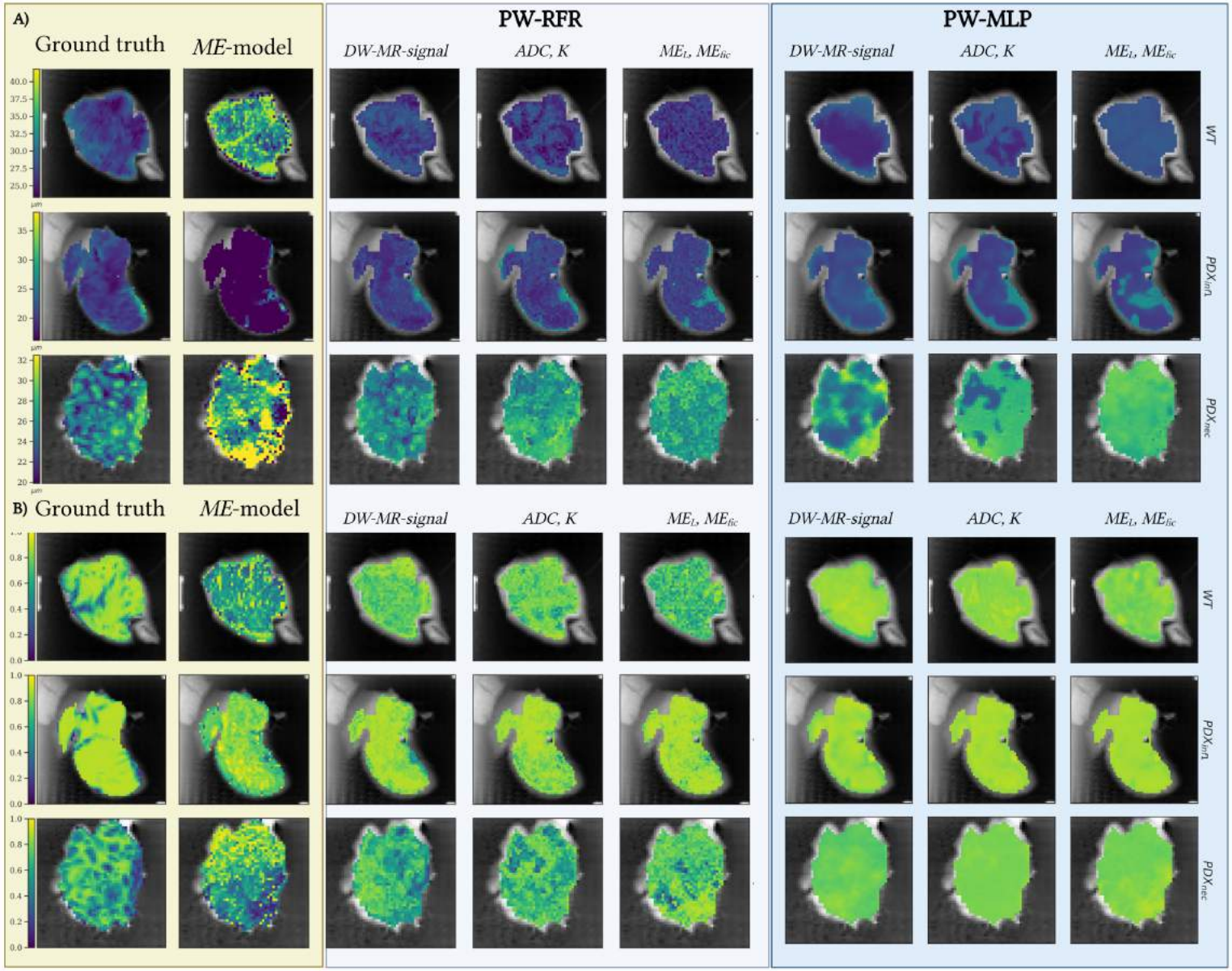


Figure 4.2: The test predictions by the PW-RFR and PW-MLP of (A) cell size L and (B) intracellular fraction f_{ic}. From left to right are visualized, the ground truth, the estimation by the ME-model, the prediction based on the DW-MR-signal, the fitted maps of ADC, K, and the parametric maps of L_{ME}, $f_{ic,ME}$ for PW-RFR and PW-MLP, respectively.

4.3.1. Algorithm comparison

In general, it appears that cell size L is slightly harder to predict than intracellular fraction f_{ic}, with for the PW-RFR based on DW-MR-signal e.g. a $REL\ MAE$ of 17.3 vs. 10.7%, respectively.

As expected, the LR does not seem to capture the complex relationship well, being a very simple algorithm. The PWLR already shows an increase in performance in comparison with LR, with the algorithms based on DW-MR-signal showing a MAE of 5.08 vs. 2.80 μm for cell size L and 0.386 vs. 0.159 for intracellular fraction f_{ic}. The KNN and RFR provide similar quantitative results.

Again, it is visible that the introduction of spatial information via the patch-wise features enhances performance, for all input types and both outcomes. The algorithms that appear to perform the best are, similar as for the simulation study, the patch-wise algorithms. Both the PW-RFR and the PW-MLP show a low error for both metrics, all input types and both outcomes. The predictions by these two algorithms on the test slice for the three representative samples are shown in Fig. 4.2. The PW-MLP, has, while close to the ground truth values on average, the tendency to predict very uniform parametric maps, mainly for f_{ic}. This limitation is especially visible for PDX_{nec}, where all contrast from the necrotic parts is lost. The PW-RFR exhibits this limitation less. For the prediction of L, the PW-MLP's parametric maps seem further off than the PW-RFR as well, showing patterns in the map that appear to much lesser degree, or not at all, in the ground truth (e.g. in the higher L part of PDX_{inf1} on the right lower side, which is enlarged for $ADC - K$, and even more for the ME-estimations). Again, this effect is somewhat present for the predictions by PW-RFR, but to a much lesser degree.

Based on these visual aspects and the similar quantitative results, it is chosen to continue the rest of the experimental approaches with the PW-RFR.

4.3.2. Input feature types

No large differences in MAE and $REL\ MAE$ are observed in the performance of the chosen model, the PW-RFR, for the three input feature types. Overall, the DW-MR-signal gives the lowest error. The prediction based on the ME-estimations is never the best, but also does not greatly underperform the other feature types, especially considering that this type has the lowest number of features, only 4 including the 2 patch-wise features computed with patch size 3×3.

When considering the predicted parametric maps in Fig. 4.2, differences are visible for the feature types, even though the quantitative results are similar. The predictions based on the ME-estimations appear to be slightly noisier than for the other two types. However, when considering the original estimation by the ME-model in the second column, a large decrease in noise is still achieved by the PW-RFR.

The algorithm is able to capture differences in cell size L (Fig. 4.2A) between the three samples, i.e. between small cell infiltration, healthy, and necrosis/inflammation, no matter the used input feature types.

Mainly in the prediction of f_{ic} for PDX_{nec} (Fig. 4.2B), differences in contrast are visible for the types. Only the algorithm based on DW-MR-signal appears to somewhat be able to capture the streak of low f_{ic} (necrosis) on the right side of the sample. In the other two predictions, on other locations regions with low f_{ic} appear. Interestingly, the prediction based on the ME-estimations appears to invert the contrast of the estimation by the ME-model, showing a high predicted f_{ic} on the lower right side vs. a low estimated $f_{ic,ME}$ and vice versa for the upper left side.

4.4. Discussion

In this experiment, we used three different input feature types, DW-MR-signal, ADC-K, and ME-estimations, to predict two microstructural characteristics of seven samples of mice livers. For each sample, a test prediction has been computed. Using these test predictions, we could compute the MAE and $REL\ MAE$ in comparison with the ground truth histology-derived maps. Based on these error metrics, it appeared that the two patch-wise algorithms, PW-RFR and PW-MLP, showed the most potential. The parametric maps as predicted by these two models were visualized for inspection, and based on that, the PW-RFR was chosen as the most promising algorithm. This algorithm will therefore be used in the next chapter, where a clinical study will be executed.

Even though the PW-MLP had the best performance for the simulation study, the predictions for the real data did not appear as accurate. In the predictions, certain patterns in the maps are observed that are not present that way in the ground truth, i.e. so called AI-hallicunations. This could be explained by the lack of sufficient and diverse training samples, further explained in the general discussion in Chapter 6.

As mentioned, the algorithms perform in general better at predicting intracellular fraction f_{ic} than they do for cell size L. This can be explained by the increased contrast that is visible in the ground truth cell size maps; the maps for f_{ic} exhibit much less variation and are more uniform, making them in theory easier to predict. The only sample for which there are large contrasts visible in f_{ic} is for PDX_{nec}. Logically, this sample appears in general to be harder to predict for the algorithms. Only the PW-RFR is able to capture some of the contrast in this sample, while the other algorithms are in general unable to capture any contrast in PDX_{nec}, predicting instead a uniform, higher, f_{ic}, and ignoring the necrotic parts with lower f_{ic}. Again, an increased diversity of the dataset with more data representative of different tissue types would possibly increase the performance to capture these contrasts (see further explanation in Chapter 6).

From the quantitative results, it appears as if even the simplest model, LR, already outperforms the fits of the ME-model. However, when looking at the individual predictions of the folds by the ME-model, in Fig. 4.2, it appears that the ME-model is able to capture the contrast between WT, i.e. healthy, and PDX_{inf1}, i.e. small cell infiltration, very well. The LR is not able to do this, and instead predicts a more uniform, average value, to get as close as possible to the true values.

Finally, the results of each algorithm per fold, i.e. per each test slice, confirm the indication that the lack of diversity in the dataset could be problematic. As stated, in principle, we have three different types of tissue available in our set. For both small cell infiltration and necrosis however, we only have one sample available. Even though we tried to circumvent this problem by ensuring there is always one representative slice available in the train set (Section 3.2.3, Fig. 3.5B), still, in general, a higher error appears especially when either PDX_{inf1} or PDX_{nec} is located in the test set. The only algorithm that appears to capture the relationship between the input features and these 'special' microstructure phenotypes is the PW-RFR, explaining the decrease in mean error over the folds as well.

4.5. Conclusion

From the experiment described in this chapter, we conclude that the PW-RFR algorithm performs the best for this application, with both high performance in the quantitive results as in the visual analysis. All input feature types give a quantitatively satisfactory performance, but show differences in the predicted test slices. The predictions based on the ME-estimations appear the most noisy, but considering the great reduction in feature dimensionality, the algorithm is still performing well. In addition, this is the

only input feature type that is independent of the input MR-acquisition protocol. Therefore, this is the feature type that will be used in the next chapter, where we will test the PW-RFR as developed on the pre-clinical data on real clinical data.

5

Clinical study

5.1. Introduction

In this chapter results are presented where the best-performing algorithm developed on pre-clinical data is deployed on *in vivo* clinical scans of patients with liver cancer. The aim of this chapter is to investigate the feasibility of using algorithms pre-trained on pre-clinical acquisitions with co-localised MRI and histology, to predict tissue microstructure in real clinical contexts.

As analysed in the previous experiment, the PW-RFR performed the best of all implemented algorithms. Only the PW-RFR trained on input feature type *(iii)* can be deployed in this clinical study. For this purpose, the algorithm will be trained on all the seven pre-clinical samples, combining the training and test set. The same algorithm settings as described in Section 3.3.1 are utilised. The trained models are exported and saved for use on the clinical data. As described, we have the parametric maps of cell size and intracellular fraction as estimated by the ME-model available. The two additional patch-wise features are formed according to the method of Section 3.2.4. The trained PW-RFR is used to predict parametric maps of cell size L and intracellular fraction f_{ic} for the three patients, based on these input features. The predictions will be used to answer *Research question 4*:

Research question 4: *Clinical study*

What is the performance of the most promising pre-trained algorithm from the pre-clinical study, when testing on clinical patient data, to predict histological cell size L and intracellular fraction f_{ic}, using parametric maps of cell size and intracellular fraction as computed by a ME-model?

In the next section, the clinical dataset is described. The adapted evaluation procedure is described in the section after that, since we cannot use the same evaluation methods as for the pre-clinical dataset. Finally, the results, a short discussion and conclusion are reported.

5.2. Methods

5.2.1. Dataset

For this clinical experiment, data from three patients participating in an ongoing, prospective phase I clinical trial on immunotherapy treatment in advanced solid tumours are used. Patient demographics and

Table 5.1: Patient demographics and primary tumour location for the three patients included in the clinical study.

Patient	Gender	Age	Primary tumour
Patient01	F	46	Melanoma
Patient02	M	55	Hepatocarcinoma
Patient03	M	55	Rectal cancer

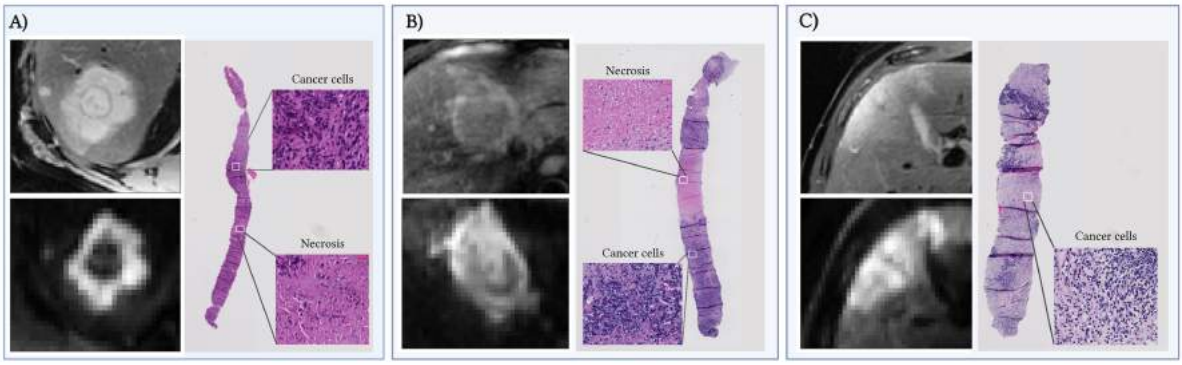


Figure 5.1: T2w and DW-MRI of the biopsied liver lesions, and the histological image of the biopsied tissue, for the three patients. On the first row, the T2w image. On the second row, the DW-image, with $b = 1500s/mm^2$ and $TE = 75ms$. Next to the MR-images, the histological images of the HE-stained biopsies. An example of the cancer cells, and necrosis if available, is shown for each patient. For A) *Patient01*, on the DWI, a clear core of necrosis is visible in the tumour (dark), surrounded by cancer cells (light). This corresponds with the findings from the biopsied tissue, containing cancer cells and areas of necrosis. For B) *Patient02*, there are cancer cells and some necrosis visible as well on the DWI. In the biopsy, two large regions of cancer cells from the tumour are visible, with an area of necrosis in between. For C) *Patient03*, no necrosis is visible, but the whole lesion exists of cancer cells. The lesion remains very bright on the DWI, while the surroundings become much darker. The biopsy contains only cancer cells, in both lumen and stroma, and no necrosis.

primary tumour locations are provided in Table 5.1. As part of the study protocol, patients underwent mpMRI with a 3T GE SIGNA Pioneer scanner along multiple time points. An ultra-sound-guided biopsy was also collected from one of the imaged liver lesions at known radiographic position after baseline MRI (i.e., before starting treatment). For this study, only the baseline scans were utilised, alongside the biopsy. The MR-protocol included, among others, anatomical T2w, T1w scans, and DW-MRI. The diffusion imaging acquisition protocol settings are the following: a slice thickness of $6mm$ and resolution of $2.4mm \times 2.4mm$, $TR = 3500ms$, $b = [0, 50, 100, 400, 900, 1200, 1500]\ s/mm^2$, and each b-value being acquired with $TE = \{75, 90, 105\}$ ms. The diffusion gradient duration and separation (δ and Δ) were: $\delta = \{0.000,\ 3.900,\ 5.200,\ 9.200,\ 15.000,\ 18.200,\ 21.000\}$ ms and $\Delta = \{0.000,\ 27.800,\ 29.000,\ 33.000,\ 28.700,\ 31.800,\ 34.700\}$ ms for $TE = 75$ ms; $\delta = \{0.000,\ 3.900,\ 5.200,\ 9.200,\ 13.000,\ 15.800,\ 18.500\}$ ms and $\Delta = \{0.000,\ 27.800,\ 29.000,\ 33.000,\ 37.000,\ 39.600,\ 42.300\}$ ms for $TE = \{90, 105\}$ ms. Examples of anatomical T2w-images and diffusion images of the biopsied liver lesions are shown in Fig. 5.1.

Unlike for the *ex vivo* pre-clinical data, this mpMRI data was acquired on patients *in vivo*. Spatially matched histological imaging is not possible for this study, and we do not have a voxel-wise ground truth available for the target. However, an ultra-sound guided biopsy was taken from the lesion imaged with mpMRI, by an experienced radiologist. The location of the biopsied lesion was segmented on the MRI. The histological material was processed similarly as the pre-clinical material, using paraffin-embedding and HE-staining, and digitized using a Hamamatsu C9600-12 scanner (resolution: $0.45\mu m$). An experienced pathologist visually inspected the images and outlined areas containing either fibrosis, necrosis, or tumour cells. Additionally, any other tissue types from outside the biopsied lesions were

outlined. The histological images of the biopsies of all three patients are visualized in Fig. 5.1, indicating examples of the outlined tissue type areas.

Furthermore, based on these histological images, cell size L and intracellular fraction f_{ic} could be computed using the same procedure in QuPath [48] as described in Section 3.2.1. The only difference is the scaling factors used to account for tissue shrinkage in the histology. For the *in vivo* data, we use the factor $1.1806\times$ to account for the histological preparation, and the factor $1.273\times$ to account for the bias induced by the 2D sectioning.

Based on the mpMRI, the same ME-model as described in Section 4.2.1, is fitted to the DW-MR-signal data (based on Eq. (4.2)). Where for the pre-clinical *ex vivo* data, only acquisitions with $b > 1000s/mm^2$ were used to ensure removal of vascular signal, we set this threshold at $b > 100s/mm^2$ for this *in vivo* study. As there is pulsation in this setting, the $ADC_V \gg 10\frac{\mu m^2}{ms}$, i.e. much higher (due to perfusion effects), and we can set the b-value threshold thus lower. The same custom-written Python script is used for the maximum likelihood fitting of the model to the MR-data. This results in parametric maps with an estimation of cell size and intracellular fraction, i.e. input feature type *(iii)* as described in Section 4.2.1.

5.2.2. Evaluation

We cannot compare the predictions voxel-by-voxel to the available 'ground truth' histological images. Instead, the mean L and f_{ic} are computed over the whole biopsy. We have the segmentations of the biopsied lesions available on the DW-MR-images. These can be used as a mask for the parametric maps of the ME-estimations, so that we know which voxels belong to the biopsied lesion, and thus for which voxels the prediction should be analysed. To be able to globally compare the predicted values with the values as computed over the biopsy, we compute the mean predicted L and f_{ic} over all the values within this *biopsy-mask*. To compare this performance with the estimation of the ME-model itself, the mean estimated L_{ME} and $f_{ic,ME}$ are also computed.

5.3. Results

5.3.1. Visual comparison

As visible from the colour bars in Fig. 5.2, the range of predicted cell size values is much smaller for the PW-RFR than for the ME-model. No clear patterns are visible in the cell size maps, except for *Patient02*. From the biopsy, we know that for *Patient02*, there is a core of necrosis in the lesion, surrounded by tumour cells. This contrast is visible in the prediction by the PW-RFR; showing lower f_{ic} in the centre of the lesion and higher f_{ic} around the borders. In the estimation of $f_{ic,ME}$, this is not captured. The same holds for *Patient01*, where this contrast is captured slightly better by the ME-model than the PW-RFR. However, also in the prediction by PW-RFR, lower f_{ic} in the centre of the lesion can be detected.

5.3.2. Quantitative comparison

In addition to the visual inspection of the parametric maps, we computed the average cell size and intracellular fraction for the ME-estimations, the PW-RFR predictions, and the biopsy. These results are visualized in Table 5.2. For all three samples the PW-RFR predicts on average a cell size L quite close to the average cell size L as found in the biopsy. Where the ME-model overestimates the cell size for each patient, the PW-RFR is able to correct this quite well. For intracellular fraction, the ME-model estimates values slightly closer to the biopsy for *Patient01* and *03*. *Patient02* has a much lower

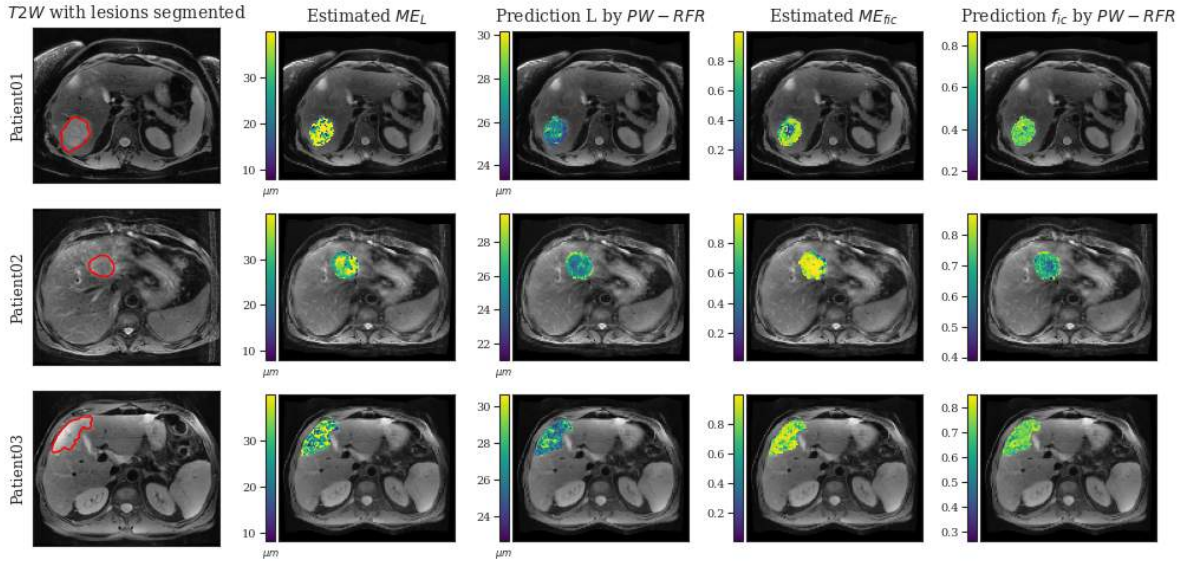


Figure 5.2: The results of testing the PW-RFR on three patients, using the ME-estimations as input features. From top to bottom: the three different patients. From left to right: a high quality T2-weighted anatomical scan, with the biopsied lesion segmented, the estimated L_{ME}, the prediction of L by PW-RFR, the estimated $f_{ic,ME}$, and the prediction of f_{ic} by PW-RFR.

f_{ic} than the others on the biopsied tissue. As was also visible from the visualization in Fig. 5.2, the PW-RFR is able to capture the lower f_{ic} (especially in the centre) better than the ME-model. This explains that the prediction is on average closer to the biopsy values than that the estimation is.

5.4. Discussion

In this experiment, we attempted to predict parametric maps of microstructural characteristics, in a clinical *in vivo* setting, using an algorithm trained on pre-clinical *ex vivo* data.

As can be seen in Table 5.2, we have found that the predictions are in a realistic range in comparison with histology-derived microstructure values obtained via biopsy of the lesions. Especially, we were able to predict a cell size very close to the histology-derived one, with biopsy vs. PW-RFR: 27.3 vs. 26.4, 25.5 vs. 25.7, 25.7 vs. $26.6\mu m$ for the three patients. This indicates that it is possible to pre-train a

Table 5.2: The mean cell size L and intracellular fraction f_{ic} as computed over the histological image obtained from tissue obtained via biopsy, as estimated by the ME-model from DW-MR-signal data, and as predicted from the parametric maps of ME_L and ME_{fic} using a PW-RFR.

	Cell size L [μm]		
	Biopsy	ME-model	PW-RFR
Patient01	27.3	31.7	26.4
Patient02	25.5	32.3	25.7
Patient03	25.7	29.1	26.6
	Intracellular fraction f_{ic} [-]		
	Biopsy	ME-model	PW-RFR
Patient01	0.842	0.733	0.711
Patient02	0.445	0.857	0.678
Patient03	0.845	0.807	0.709

model on pre-clinical data, and consequently use it on real patient data, discussed further in the next chapter (Chapter 6).

As shown in Fig. 5.1, both the lesions of *Patient01* and *02* show necrosis in the core of the tumour, surrounded by cancer cells. Especially in the biopsy of *Patient02* (Fig. 5.1), this is clearly visible, also corresponding with the low ($f_{ic} = 0.445$, the lowest of all patients) as computed from the biopsy. While in the estimation by the ME-model this is not replicated (0.857, the highest of all patients), the PW-RFR's prediction improves this greatly (0.678, the lowest of all patients). The same improvement is visible in the predicted maps in Fig. 5.2, showing a clear lowered f_{ic} in the core of the lesion.

While these results look promising, this is still a very preliminary study. Where we were able to execute a voxel-by-voxel comparison of the whole parametric map for the pre-clinical study, we could now only compare the absolute values with the biopsy. As described in the Section 2.1.1, only a very small part of the lesion is sampled with this method, of which we do not know the exact location. Therefore, we compared the average predicted values over the whole lesion with those of the biopsy. Due to the outlining on the histological imaging by the pathologist, we know what kind of microstructure types are samples in the biopsy, e.g. necrosis, or cancer cells, and we can assume that the biopsied tissue is representative for the whole lesion. However, as we do not have a complete map available of the microstructural characteristics, the conclusions we draw about the performance of the algorithm remain speculative.

5.5. Conclusion

Despite these limitations, our results suggest that it may be feasible to use co-localised MRI-histology data obtained pre-clinically to train AI-algorithms that predict microstructure from DW-MRI, and then deploy them on clinical scans obtained in vivo. Nonetheless, further work will be needed to fully assess the performance of such algorithms and validate their predictions on histological data.

6

General Discussion

The main research objective of this thesis was:

Main research objective

The development of a novel AI-based method to estimate microstructural parameters in the liver, specifically cell size L and intracellular fraction f_{ic}, from DW-MRI data, outperforming the estimation of a biophysical model.

The main research aim was pursued via three experiments. Firstly, a simulation study was executed to implement the algorithms, using data for which a clear relationship between the input features X and the targets y was known. Seven algorithms were used, of which four voxel-wise ML (LR, PWLR, KNN, RFR), one patch-wise ML (PW-RFR), one voxel-wise DL (MLP), and one patch-wise DL (PW-MLP). From the simulation study, it could already be concluded that the incorporation of patch-wise features enhances the performance of the algorithms greatly, with the PW-RFR and the PW-MLP showing the lowest errors, both absolutely and relatively.

In the next experiment, real pre-clinical data was used to answer *Research questions 2* and *3*. Based on quantitive evaluation metrics and visual inspections of the predictions in comparison with the histology-derived ground truth, the PW-RFR was found to be the most promising algorithm, in correspondence with the results from the simulation study. Furthermore, no large differences were observed in the performance of the chosen algorithm for the three input feature types, exhibiting only small differences in the evaluation metrics. The predictions based on the DW-MR-signal replicate the contrasts from the ground truth slightly better. However, this input feature type is also the most protocol-dependant, the least interpretable, and has the highest dimensionality, resulting in the most complex model. The use of either the fitted $ADC - K$ or parametric maps of the ME-model causes a reduction in feature dimensionality.

For our clinical study described in the third experiment however, we were only able to use the last input feature type; the ME-estimations. This was due to the fact that the pre-clinical and clinical MR-protocol are different, and this is the only input feature type with features (parametric maps of L_{ME} and $f_{ic,ME}$) independent of the input MR-protocol. In this experiment, we exported the PW-RFR, pre-trained on the pre-clinical data with this input feature type *(iii)*. This PW-RFR was consequently

deployed on the data of three patients with liver lesions. For these lesions, there was also an ultrasound guided biopsy available. The predicted maps of the microstructural characteristics were visually analysed, and compared with quantitative data obtained via histological images of the biopsied tissue. The ability to train an algorithm on pre-clinical data and subsequently apply it to clinical data presents a notable advantage in this study. Spatially-matched data like we have is scarce, and would be even more challenging to acquire from a clinical setting. Therefore, it is beneficial to leverage a pre-trained model trained on spatially matched pre-clinical data, allowing for predictions on clinical data without the need of a histological ground truth. The voxel-by-voxel comparison of spatially matched MRI and histology maps remains the ideal method for exhaustive histological validation studies, considering there might be differences in the relationship of the observed DW-MR-signal and microstructure, between pre-clinical and clinical settings. However, the application of pre-trained models derived from pre-clinical data to clinical data provides valuable insights, bridging the gap between pre-clinical research and clinical applications. The results of our clinical study are, while promising, preliminary, and further validation and verification on larger and more diverse clinical datasets would be necessary to confirm the algorithm's reliability in clinical contexts.

A number of potential limitations are acknowledged for this thesis and will be discussed in the following paragraphs. Firstly, the cell size and intracellular fraction as derived from histological images of the tissue serve as the ground truth for this study. However, using this histological ground truth itself has certain inherent limitations, specifically arising due to the fact that we compare small 2D sections of histological material to 3D MRI data. As mentioned, there is a mismatch in spatial scales between MRI and histology. The MRI slices used in this study have a thickness ranging of $2.2mm$, whereas the histology sections are much thinner, $4\,\mu m$. As a result, portions of the tissue that contribute to the measured DW-MR signals may not be included in the histological maps due to the limitations of histological sectioning. This discrepancy in spatial resolution raises the possibility that certain within-sample variations observed in the MRI data are not directly reflected in the histology, limiting the ability to precisely predict the observed microstructural characteristics. The correspondence between the microstructural characteristics derived from histology and those predicted from DW-MRI data thus has some degree of variability and uncertainty. This should be considered when interpreting the findings of this study, as the used histological ground truth, although valuable, may not fully correspond to and capture the microstructure as imaged on the MRI.

Secondly, as mentioned, the number of samples available for our analyses was limited, despite being higher than in most studies where co-registered MRI-histology data are used. With no more than seven samples at hand, it would be challenging to properly train algorithms. To mitigate this problem, we employ a voxel-wise training approach, resulting in ~ 15.000 independent training 'samples'. While this strategy increases the effective number of training samples, it does not address the underlying issue of limited diversity within the dataset. Visual inspection of the mouse liver histology by an experienced pathologist reveals that in practical terms, the 7 samples contained only 3 unique types of microstructure, namely: i) normal liver structures; ii) infiltration of small malignant cells in between hepatocytes (lymphoma-like); iii) extensive necrosis and inflammation. These are clearly represented in samples WT (normal liver), PDX_{inf1} (small cell infiltration), and PDX_{nec} (necrosis and inflammation), as shown in Figure 3.1 in Chapter 3. The limited number of microstructures available may have hampered the algorithm's ability to generalize and accurately predict the full spectrum of microstructural characteristics. Future studies should aim to obtain a more diverse and representative dataset, to enhance the robustness and effectiveness of the predictive models in capturing the full range of microstructural characteristics.

In an attempt to improve the data diversity, we decided to slightly adapt the conventional LOO-CV (see Fig. 3.5 as described in Section 3.2.3), by moving one slice of the test set in each fold back to the training set, while the one remaining slice was now treated as the test sample. However, this approach introduces a potential risk of data leakage, since information from the test sample, i.e. the one slice, is included in the training set. As a result, the algorithms may inadvertently learn and exploit this leaked information during training, leading to an overestimation of the predictive performance during cross-validation in the pre-clinical study.

Interestingly, the algorithms seem to be affected in various ways by the remaining scarcity of diversity. Despite this limitation, mainly the PW-RFR appears to exhibit a good performance in predicting also the underrepresented microstructure types. While the PW-MLP showed a good performance for the simulation study, the results on the real pre-clinical data were less promising. This observation suggests that the PW-RFR algorithm might possess greater resilience to the issue of the limited diversity. While both algorithms are naturally affected, the MLP's more complex architecture makes it particularly susceptible to the lack of diversity. Consequently, the trained network may struggle to capture the nuanced features and patterns related to specialized microstructures, potentially leading to overfitting and the observed exaggeration of certain features that are not present in that way in the output (the so called AI-hallicunations, as mentioned in Chapter 4). The (PW-)RFR's reliance on explicit feature engineering and ensemble learning mechanisms may contribute to its relatively better performance. This finding emphasizes the importance of dataset diversity and suggests that, while the RFR may be a more suitable choice for predicting microstructural characteristics from DW-MRI data for this specific dataset, further investigations are needed to validate and generalize these findings across different datasets.

In the context of predicting parametric maps from MR-images [29], convolutional neural networks (CNNs) are often considered the go-to approach due to their effectiveness in capturing spatial dependencies within images. However, in this particular study, utilizing a CNN for prediction becomes impractical due to the adopted voxel-wise training approach. Typically, CNNs require the entire image as input for training, which would result in an extremely small training set size given the limited number of seven samples available. Such a small training set poses a significant challenge for CNNs, as they rely on a substantial amount of data to learn and generalize complex spatial patterns effectively. Consequently, the voxel-wise approach was employed to augment the effective training set size and overcome the limitation of limited data. The large disadvantage is of this approach is that all spatial information is lost, treating each voxel totally independent. Naturally, however, a relationship exists between the microstructure of neighbouring voxels and incorporating this could improve the performance of the prediction [56]. In an attempt to do this, we included patch-wise features for two of the algorithms as an alternative approach to use spatial information. The finding that these algorithms showed the highest performance supports the hypothesis that a CNN might be a suitable algorithm for this application, granting availability of a sufficient amount of training data.

Finally, another limitation of this study relates to the method used for hyperparameter optimization in the implementation of the algorithms. The simulated dataset was employed to explore and evaluate various hyperparameter configurations for some of the algorithms. However, it is important to note that this hyperparameter optimization process was not conducted extensively or exhaustively. Instead of performing a comprehensive grid search or systematically exploring the entire hyperparameter space, a limited set of hyperparameters was tested, and the optimal hyperparameters were chosen based on their performance on the simulated dataset. These chosen hyperparameter configurations were then utilized for the algorithms' application on the real pre-clinical and clinical dataset. Consequently, this

approach to hyperparameter optimization may not fully capture the optimal settings for each algorithm on the real data. Potentially, better hyperparameter configurations exist, which could improve the predictive performance. The performance of the algorithms in this study can therefore be considered as a so-called baseline performance. Future studies should consider employing more rigorous and systematic techniques for hyperparameter optimization, such as grid search or more advanced methods like Bayesian optimization or random search, to thoroughly explore the hyperparameter space and identify the most suitable configurations. Finally, this is not only a limitation for the hyperparameters of the DNN, but also for the chosen architecture in general. We now utilise a relatively simple MLP-architecture, leaving many promising architectures unexplored, such as Recurrent Neural Networks (RNNs) and Transformers. The possibility that the use of these architectures could result in a better performance is present, but is out of the scope of this thesis.

6.1. Conclusion

In summary, this thesis aimed to develop an AI-based method for estimating microstructural parameters from DW-MRI data. The PW-RFR algorithm demonstrated a promising performance in both simulated and real pre-clinical data, outperforming other algorithms. The ability to deploy models that have been trained on pre-clinical data to real-world clinical contexts provided valuable insights, bridging the gap between research and clinical applications. It provided promising results pointing towards the feasibility of developing tools that learn how to predict histopathological characteristics from DW-MRI on large databases of pre-clinical MRI-histology data. However, the results of the clinical study presented in this thesis are preliminary, and further validation on larger and diverse datasets would be required. Limitations include spatial resolution mismatch between MRI and histology, limited data availability (specifically in diversity of microstructure types), and the need for more extensive hyperparameter optimization. Future research should focus on addressing these limitations and exploring advanced DL architectures such as CNNs, RNNs, and transformers. Overall, this thesis provides insights into the development of AI-based methods for liver microstructural parameter estimation based on DW-MRI.

References

1. Martinez-Heras E, Grussu F, Prados F, Solana E, and Llufriu S. Diffusion-Weighted Imaging: Recent Advances and Applications. Seminars in Ultrasound, CT and MRI 2021 Oct; 42:490–506. DOI: 10.1053/J.SULT.2021.07.006
2. Novikov DS, Fieremans E, Jespersen SN, and Kiselev VG. Quantifying brain microstructure with diffusion MRI: Theory and parameter estimation. NMR in Biomedicine 2019 Apr; 32:e3998. DOI: 10.1002/NBM.3998. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/nbm.3998%20https://onlinelibrary.wiley.com/doi/abs/10.1002/nbm.3998%20https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/nbm.3998
3. Alexander DC, Dyrby TB, Nilsson M, and Zhang H. Imaging brain microstructure with diffusion MRI: practicality and applications. NMR in Biomedicine 2019 Apr; 32:e3841. DOI: 10.1002/NBM.3841. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/nbm.3841%20https://onlinelibrary.wiley.com/doi/abs/10.1002/nbm.3841%20https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/nbm.3841
4. Einstein A. On the motion of small particles suspended in liquids at rest required by the molecular-kinetic theory of heat. Annalen der physik 1905; 17:208
5. Reynaud O. Time-dependent diffusion MRI in cancer: Tissue modeling and applications. Frontiers in Physics 2017 Nov; 5:58. DOI: 10.3389/FPHY.2017.00058/BIBTEX
6. Kiselev VG. Microstructure with diffusion MRI: what scale we are sensitive to? Journal of Neuroscience Methods 2021; 347:108910. DOI: 10.1016/j.jneumeth.2020.108910. Available from: https://doi.org/10.1016/j.jneumeth.2020.108910
7. Chatterjee A, Harmath C, and Oto A. New prostate MRI techniques and sequences. Abdominal Radiology 2020 Dec; 45:4052–62. DOI: 10.1007/S00261-020-02504-8/TABLES/1. Available from: https://link.springer.com/article/10.1007/s00261-020-02504-8
8. Salo RA, Belevich I, Jokitalo E, Gröhn O, and Sierra A. Assessment of the structural complexity of diffusion MRI voxels using 3D electron microscopy in the rat brain. NeuroImage 2021 Jan; 225:117529. DOI: 10.1016/J.NEUROIMAGE.2020.117529
9. Dyrby TB, Innocenti GM, Bech M, and Lundell H. Validation strategies for the interpretation of microstructure imaging using diffusion MRI. NeuroImage 2018 Nov; 182:62–79. DOI: 10.1016/J.NEUROIMAGE.2018.06.049
10. Jelescu IO, Veraart J, Fieremans E, and Novikov DS. Degeneracy in model parameter estimation for multi-compartmental diffusion in neuronal tissue. NMR in Biomedicine 2016 Jan; 29:33–47. DOI: 10.1002/nbm.3450
11. Harms RL, Fritz FJ, Tobisch A, Goebel R, and Roebroeck A. Robust and fast nonlinear optimization of diffusion MRI microstructure models. NeuroImage 2017; 155:82–96. DOI: 10.1016/j.neuroimage.2017.04.064. Available from: http://dx.doi.org/10.1016/j.neuroimage.2017.04.064

12. Ferizi U, Schneider T, Panagiotaki E, Nedjati-Gilani G, Zhang H, Wheeler-Kingshott CA, and Alexander DC. A ranking of diffusion MRI compartment models with in vivo human brain data. Magnetic Resonance in Medicine 2014 Dec; 72:1785–92. DOI: 10.1002/MRM.25080/ASSET/SUPINFO/MRM25080-SUP-0002-SUPPINFO2.DOCX. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/mrm.25080%20https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.25080%20https://onlinelibrary.wiley.com/doi/10.1002/mrm.25080

13. Hutchinson EB, Avram AV, Irfanoglu MO, Koay CG, Barnett AS, Komlosh ME, Özarslan E, Schwerin SC, Juliano SL, and Pierpaoli C. Analysis of the effects of noise, DWI sampling, and value of assumed parameters in diffusion MRI models. Magnetic Resonance in Medicine 2017 Nov; 78:1767–80. DOI: 10.1002/MRM.26575/ASSET/SUPINFO/MRM26575-SUP-0001-SUPPINFO.PDF. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/mrm.26575%20https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.26575%20https://onlinelibrary.wiley.com/doi/10.1002/mrm.26575

14. Rokem A, Yeatman JD, Pestilli F, Kay KN, Mezer A, Van Der Walt S, and Wandell BA. Evaluating the Accuracy of Diffusion MRI Models in White Matter. PLOS ONE 2015 Apr; 10:e0123272. DOI: 10.1371/JOURNAL.PONE.0123272. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0123272

15. Minati L and Weglarz WP. Physical foundations, models, and methods of diffusion magnetic resonance imaging of the brain: A review. Concepts in Magnetic Resonance Part A 2007 Sep; 30A:278–307. DOI: 10.1002/CMR.A.20094. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/cmr.a.20094%20https://onlinelibrary.wiley.com/doi/abs/10.1002/cmr.a.20094%20https://onlinelibrary.wiley.com/doi/10.1002/cmr.a.20094

16. Jelescu IO and Budde MD. Design and validation of diffusion MRI models of white matter. Frontiers in Physics 2017 Nov; 5:61. DOI: 10.3389/FPHY.2017.00061/BIBTEX

17. Tang L and Zhou XJ. Diffusion MRI of cancer: From low to high b-values. 2019 Jan. DOI: 10.1002/jmri.26293

18. Brancato V, Cavaliere C, Salvatore M, and Monti S. Non-Gaussian models of diffusion weighted imaging for detection and characterization of prostate cancer: a systematic review and meta-analysis. Scientific Reports 2019 9:1 2019 Nov; 9:1–15. DOI: 10.1038/s41598-019-53350-8. Available from: https://www.nature.com/articles/s41598-019-53350-8

19. Galbán CJ, Hoff BA, Chenevert TL, and Ross BD. Diffusion MRI in early cancer therapeutic response assessment. NMR in biomedicine 2017 Mar; 30. DOI: 10.1002/NBM.3458. Available from: https://pubmed.ncbi.nlm.nih.gov/26773848/

20. Lanzman RS and Wittsack HJ. Diffusion tensor imaging in abdominal organs. NMR in Biomedicine 2017 Mar; 30:e3434. DOI: 10.1002/NBM.3434. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/nbm.3434%20https://onlinelibrary.wiley.com/doi/abs/10.1002/nbm.3434%20https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/nbm.3434

21. Henriques RN, Jespersen SN, and Shemesh N. Evidence for microscopic kurtosis in neural tissue revealed by correlation tensor MRI. Magnetic Resonance in Medicine 2021 Dec; 86:3111–30. DOI: 10.1002/mrm.28938

22. Surov A, Meyer HJ, and Wienke A. Correlation between apparent diffusion coefficient (ADC) and cellularity is different in several tumors: a meta-analysis. Oncotarget 2017; 8:59492–9. DOI: 10.18632/ONCOTARGET.17752. Available from: https://pubmed.ncbi.nlm.nih.gov/28938652/

23. Perez-Lopez R, Mateo J, Mossop H, Blackledge MD, Collins DJ, Rata M, Morgan VA, Macdonald A, Sandhu S, Lorente D, Rescigno P, Zafeiriou Z, Bianchini D, Porta N, Hall E, Leach MO, Bono JS de, Koh DM, and Tunariu N. Diffusion-weighted Imaging as a Treatment Response Biomarker for Evaluating Bone Metastases in Prostate Cancer: A Pilot Study. Radiology 2017 Apr; 283:168–77. DOI: 10.1148/radiol.2016160646

24. Panagiotaki E, Chan RW, Dikaios N, Ahmed HU, O'Callaghan J, Freeman A, Atkinson D, Punwani S, Hawkes DJ, and Alexander DC. Microstructural characterization of normal and malignant human prostate tissue with vascular, extracellular, and restricted diffusion for cytometry in tumours magnetic resonance imaging. Investigative radiology 2015 Apr; 50:218–27. DOI: 10.1097/RLI.0000000000000115. Available from: https://pubmed.ncbi.nlm.nih.gov/25426656/

25. Grussu F, Bernatowicz K, Casanova-Salas I, Castro N, Nuciforo P, Mateo J, Barba I, and Perez-Lopez R. Diffusion MRI signal cumulants and hepatocyte microstructure at fixed diffusion time: Insights from simulations, 9.4T imaging, and histology. Magnetic Resonance in Medicine 2022 Jul; 88:365–79. DOI: 10.1002/MRM.29174. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/mrm.29174%20https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.29174%20https://onlinelibrary.wiley.com/doi/10.1002/mrm.29174

26. Ianuş A, Drobnjak I, and Alexander DC. Model-based estimation of microscopic anisotropy using diffusion MRI: a simulation study. NMR in Biomedicine 2016 May; 29:672–85. DOI: 10.1002/NBM.3496. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/nbm.3496%20https://onlinelibrary.wiley.com/doi/abs/10.1002/nbm.3496%20https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/nbm.3496

27. Panagiotaki E, Walker-Samuel S, Siow B, Johnson SP, Rajkumar V, Pedley RB, Lythgoe MF, and Alexander DC. Noninvasive quantification of solid tumor microstructure using VERDICT MRI. Cancer Research 2014; 74:1902–12. DOI: 10.1158/0008-5472.CAN-13-2511

28. Jiang X, Devan SP, Xie J, Gore JC, and Xu J. Improving MR cell size imaging by inclusion of transcytolemmal water exchange. NMR in biomedicine 2022. DOI: 10.1002/NBM.4799. Available from: https://pubmed.ncbi.nlm.nih.gov/35794795/

29. To MNN and Kwak JT. Biparametric MR signal characteristics can predict histopathological measures of prostate cancer. European Radiology 2022. DOI: 10.1007/s00330-022-08808-1

30. Sun Y, Reynolds HM, Wraith D, Williams S, Finnegan ME, Mitchell C, Murphy D, and Haworth A. Voxel-wise prostate cell density prediction using multiparametric magnetic resonance imaging and machine learning. Acta Oncologica 2018 Dec; 57:1540–6. DOI: 10.1080/0284186X.2018.1468084. Available from: https://www.tandfonline.com/action/journalInformation?journalCode=ionc20

31. Brabec J, Friedjungová M, Vašata D, Englund E, Bengzon J, Knutsson L, Szczepankiewicz F, Westen D van, Sundgren PC, and Nilsson M. Meningioma microstructure assessed by diffusion MRI: An investigation of the source of mean diffusivity and fractional anisotropy by quantitative histology. NeuroImage: Clinical 2023 Jan; 37:103365. DOI: 10.1016/J.NICL.2023.103365

32. Padhani AR, Liu G, Mu-Koh D, Chenevert TL, Thoeny HC, Takahara T, Dzik-Jurasz A, Ross BD, Van Cauteren M, Collins D, Hammoud DA, Rustin GJ, Taouli B, and Choyke PL. Diffusion-Weighted Magnetic Resonance Imaging as a Cancer Biomarker: Consensus and Recommendations. Neoplasia (New York, N.Y.) 2009; 11:102. DOI: 10.1593/NEO.81328. Available from: /pmc/articles/PMC2631136/%20/pmc/articles/PMC2631136/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2631136/

33. Xu J, Jiang X, Devan SP, Arlinghaus LR, McKinley ET, Xie J, Zu Z, Wang Q, Chakravarthy AB, Wang Y, and Gore JC. MRI-Cytometry: Mapping non-parametric cell size distributions using diffusion MRI. Magnetic resonance in medicine 2021 Feb; 85:748. DOI: 10.1002/MRM.28454. Available from: /pmc/articles/PMC7722100/%20/pmc/articles/PMC7722100/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7722100/

34. Komuta M. Histological Heterogeneity of Primary Liver Cancers: Clinical Relevance, Diagnostic Pitfalls and the Pathologist's Role. Cancers 2021, Vol. 13, Page 2871 2021 Jun; 13:2871. DOI: 10.3390/CANCERS13122871. Available from: https://www.mdpi.com/2072-6694/13/12/2871/htm%20https://www.mdpi.com/2072-6694/13/12/2871

35. Roberts TA, Hyare H, Agliardi G, Hipwell B, d'Esposito A, Ianus A, Breen-Norris JO, Ramasawmy R, Taylor V, Atkinson D, Punwani S, Lythgoe MF, Siow B, Brandner S, Rees J, Panagiotaki E, Alexander DC, and Walker-Samuel S. Noninvasive diffusion magnetic resonance imaging of brain tumour cell size for the early detection of therapeutic response. Scientific Reports 2020 10:1 2020 Jun; 10:1–13. DOI: 10.1038/s41598-020-65956-4. Available from: https://www.nature.com/articles/s41598-020-65956-4

36. Tobkes AI and Nord HJ. Liver Biopsy: Review of Methodology and Complications. Digestive Diseases 1995; 13:267–74. DOI: 10.1159/000171507. Available from: https://www.karger.com/Article/FullText/171507%20https://www.karger.com/Article/Abstract/171507

37. Friedman LS. Controversies in liver biopsy: Who, where, when, how, why? Current Gastroenterology Reports 2004; 6:30–6. DOI: 10.1007/S11894-004-0023-4/METRICS. Available from: https://link.springer.com/article/10.1007/s11894-004-0023-4

38. Shyamala K, Girish H, and Murgod S. Risk of tumor cell seeding through biopsy and aspiration cytology. Journal of International Society of Preventive & Community Dentistry 2014 Jan; 4:5. DOI: 10.4103/2231-0762.129446. Available from: /pmc/articles/PMC4015162/%20/pmc/articles/PMC4015162/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4015162/

39. Nilsson M, Englund E, Szczepankiewicz F, Westen D van, and Sundgren PC. Imaging brain tumour microstructure. NeuroImage 2018 Nov; 182:232–50. DOI: 10.1016/J.NEUROIMAGE.2018.04.075

40. Palombo M, Valindria V, Singh S, Chiou E, Giganti F, Pye H, Whitaker HC, Atkinson D, Punwani S, Alexander DC, and Panagiotaki E. Joint estimation of relaxation and diffusion tissue parameters for prostate cancer with relaxation-VERDICT MRI. Scientific Reports 2023 13:1 2023 Feb; 13:1–13. DOI: 10.1038/s41598-023-30182-1. Available from: https://www.nature.com/articles/s41598-023-30182-1

41. Bourne RM, Bailey C, Johnston EW, Pye H, Heavey S, Whitaker H, Siow B, Freeman A, Shaw GL, Sridhar A, Mertzanidou T, Hawkes DJ, Alexander DC, Punwani S, and Panagiotaki E. Apparatus for histological validation of in vivo and ex vivo magnetic resonance imaging of the human prostate. Frontiers in Oncology 2017 Mar; 7:47. DOI: 10.3389/FONC.2017.00047/BIBTEX

42. Stejskal EO and Tanner JE. Spin Diffusion Measurements: Spin Echoes in the Presence of a Time□Dependent Field Gradient. The Journal of Chemical Physics 1965 Jul; 42:288. DOI: 10.1063/1.1695690. Available from: https://aip.scitation.org/doi/abs/10.1063/1.1695690

43. Fieremans E and Lee HH. Physical and numerical phantoms for the validation of brain microstructural MRI: A cookbook. NeuroImage 2018 Nov; 182:39–61. DOI: 10.1016/J.NEUROIMAGE.2018.06.046. Available from: https://pubmed.ncbi.nlm.nih.gov/29920376/

44. Huisman TA. Diffusion-weighted and diffusion tensor imaging of the brain, made easy. Cancer Imaging 2010; 10:S163. DOI: 10.1102/1470-7330.2010.9023. Available from: /pmc/articles/PMC2967146/%20/pmc/articles/PMC2967146/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2967146/

45. Kiselev VG. The Cumulant Expansion: An Overarching Mathematical Framework For Understanding Diffusion NMR. *Diffusion MRI*. Oxford University Press, 2010 Nov :152–68. DOI: 10.1093/med/9780195369779.003.0010

46. Jensen JH, Helpern JA, Ramani A, Lu H, and Kaczynski K. Diffusional kurtosis imaging: The quantification of non-gaussian water diffusion by means of magnetic resonance imaging. Magnetic Resonance in Medicine 2005 Jun; 53:1432–40. DOI: 10.1002/MRM.20508. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/mrm.20508%20https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.20508%20https://onlinelibrary.wiley.com/doi/10.1002/mrm.20508

47. Lee HH, Papaioannou A, Novikov DS, and Fieremans E. In vivo observation and biophysical interpretation of time-dependent diffusion in human cortical gray matter. NeuroImage 2020 Nov; 222:117054. DOI: 10.1016/j.neuroimage.2020.117054

48. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, McQuaid S, Gray RT, Murray LJ, Coleman HG, James JA, Salto-Tellez M, and Hamilton PW. QuPath: Open source software for digital pathology image analysis. Scientific Reports 2017 7:1 2017 Dec; 7:1–7. DOI: 10.1038/s41598-017-17204-5. Available from: https://www.nature.com/articles/s41598-017-17204-5

49. Veraart J, Nunes D, Rudrapatna U, Fieremans E, Jones DK, Novikov DS, and Shemesh N. Noninvasive quantification of axon radii using diffusion MRI. eLife 2020 Feb; 9. DOI: 10.7554/ELIFE.49855

50. Garyfallidis E, Brett M, Amirbekian B, Rokem A, Walt S van der, Descoteaux M, and Nimmo-Smith I. Dipy, a library for the analysis of diffusion MRI data. Frontiers in Neuroinformatics 2014 Feb; 8:8. DOI: 10.3389/FNINF.2014.00008/BIBTEX

51. Almeida Martins JP de, Nilsson M, Lampinen B, Palombo M, While PT, Westin CF, and Szczepankiewicz F. Neural networks for parameter estimation in microstructural MRI: Application to a diffusion-relaxation model of white matter. NeuroImage 2021 Dec; 244:118601. DOI: 10.1016/J.NEUROIMAGE.2021.118601

52. Balinov B, Jönsson B, Linse P, and Söderman O. The NMR Self-Diffusion Method Applied to Restricted Diffusion. Simulation of Echo Attenuation from Molecules in Spheres and between Planes. Journal of Magnetic Resonance, Series A 1993 Aug; 104:17–25. DOI: 10.1006/JMRA.1993.1184

53. Gudbjartsson H and Patz S. The Rician Distribution of Noisy MRI Data. Magnetic resonance in medicine 1995; 34:910. DOI: 10.1002/MRM.1910340618. Available from: /pmc/articles/PMC2254141/%20/pmc/articles/PMC2254141/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2254141/

54. Yang J, Fan J, Ai D, Zhou S, Tang S, and Wang Y. Brain MR image denoising for Rician noise using pre-smooth non-local means filter. BioMedical Engineering Online 2015 Jan; 14:1–20. DOI: 10.1186/1475-925X-14-2/TABLES/3. Available from: https://biomedical-engineering-online.biomedcentral.com/articles/10.1186/1475-925X-14-2

55. Botchkarev A. Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology

56. Alexander DC, Zikic D, Ghosh A, Tanno R, Wottschel V, Zhang J, Kaden E, Dyrby TB, Sotiropoulos SN, Zhang H, and Criminisi A. Image quality transfer and applications in diffusion MRI. NeuroImage 2017 May; 152:283–98. DOI: 10.1016/J.NEUROIMAGE.2017.02.089

57. Hornik K, Stinchcombe M, and White H. Multilayer feedforward networks are universal approximators. Neural Networks 1989 Jan; 2:359–66. DOI: 10.1016/0893-6080(89)90020-8

58. Messina C, Bignone R, Bruno A, Bruno A, Bruno F, Calandri M, Caruso D, Coppolino P, De Robertis R, Gentili F, Grazzini I, Natella R, Scalise P, Barile A, Grassi R, and Albano D. Diffusion-Weighted Imaging in Oncology: An Update. Cancers 2020 Jun; 12:1–28. DOI: 10.3390/CANCERS12061493. Available from: /pmc/articles/PMC7352852/%20/pmc/articles/PMC7352852/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7352852/

59. Deen SS, Priest AN, McLean MA, Gill AB, Brodie C, Crawford R, Latimer J, Baldwin P, Earl HM, Parkinson C, Smith S, Hodgkin C, Patterson I, Addley H, Freeman S, Moyle P, Jimenez-Linan M, Graves MJ, Sala E, Brenton JD, and Gallagher FA. Diffusion kurtosis MRI as a predictive biomarker of response to neoadjuvant chemotherapy in high grade serous ovarian cancer. Scientific reports 2019 Dec; 9. DOI: 10.1038/S41598-019-47195-4. Available from: https://pubmed.ncbi.nlm.nih.gov/31341212/

60. Rosenkrantz AB, Sigmund EE, Winnick A, Niver BE, Spieler B, Morgan GR, and Hajdu CH. Assessment of hepatocellular carcinoma using apparent diffusion coefficient and diffusion kurtosis indices: preliminary experience in fresh liver explants. Magnetic resonance imaging 2012 Dec; 30:1534–40. DOI: 10.1016/J.MRI.2012.04.020. Available from: https://pubmed.ncbi.nlm.nih.gov/22819175/

61. Wu G, Zhao Z, Yao Q, Kong W, Xu J, Zhang J, Liu G, and Dai Y. The Study of Clear Cell Renal Cell Carcinoma with MR Diffusion Kurtosis Tensor Imaging and Its Histopathologic Correlation. Academic radiology 2018 Apr; 25:430–8. DOI: 10.1016/J.ACRA.2017.10.016. Available from: https://pubmed.ncbi.nlm.nih.gov/29198944/

62. Hectors SJ, Semaan S, Song C, Lewis S, Haines GK, Tewari A, Rastinehad AR, and Taouli B. Advanced Diffusion-weighted Imaging Modeling for Prostate Cancer Characterization: Correlation with Quantitative Histopathologic Tumor Tissue Composition-A Hypothesis-generating Study. Radiology 2018 Mar; 286:918–28. DOI: 10.1148/RADIOL.2017170904. Available from: https://pubmed.ncbi.nlm.nih.gov/29117481/

63. Guo J, Dong C, Wu Z, Sun W, Li X, Zhou R, and Xu W. Diffusion kurtosis imaging assessment of the response to radiotherapy in a VX2 bone tumor model: an animal study. Acta radiologica (Stockholm, Sweden : 1987) 2022 Feb; 63:182–91. DOI: 10.1177/0284185121989519. Available from: https://pubmed.ncbi.nlm.nih.gov/33535770/

64. Grussu F, Battiston M, Veraart J, Schneider T, Cohen-Adad J, Shepherd TM, Alexander DC, Fieremans E, Novikov DS, and Gandini Wheeler-Kingshott CA. Multi-parametric quantitative in vivo spinal cord MRI with unified signal readout and image denoising. NeuroImage 2020 Aug; 217. DOI: 10.1016/J.NEUROIMAGE.2020.116884. Available from: https://pubmed.ncbi.nlm.nih.gov/32360689/

A

Literature Review

See additional material.

B

Results of the simulation study

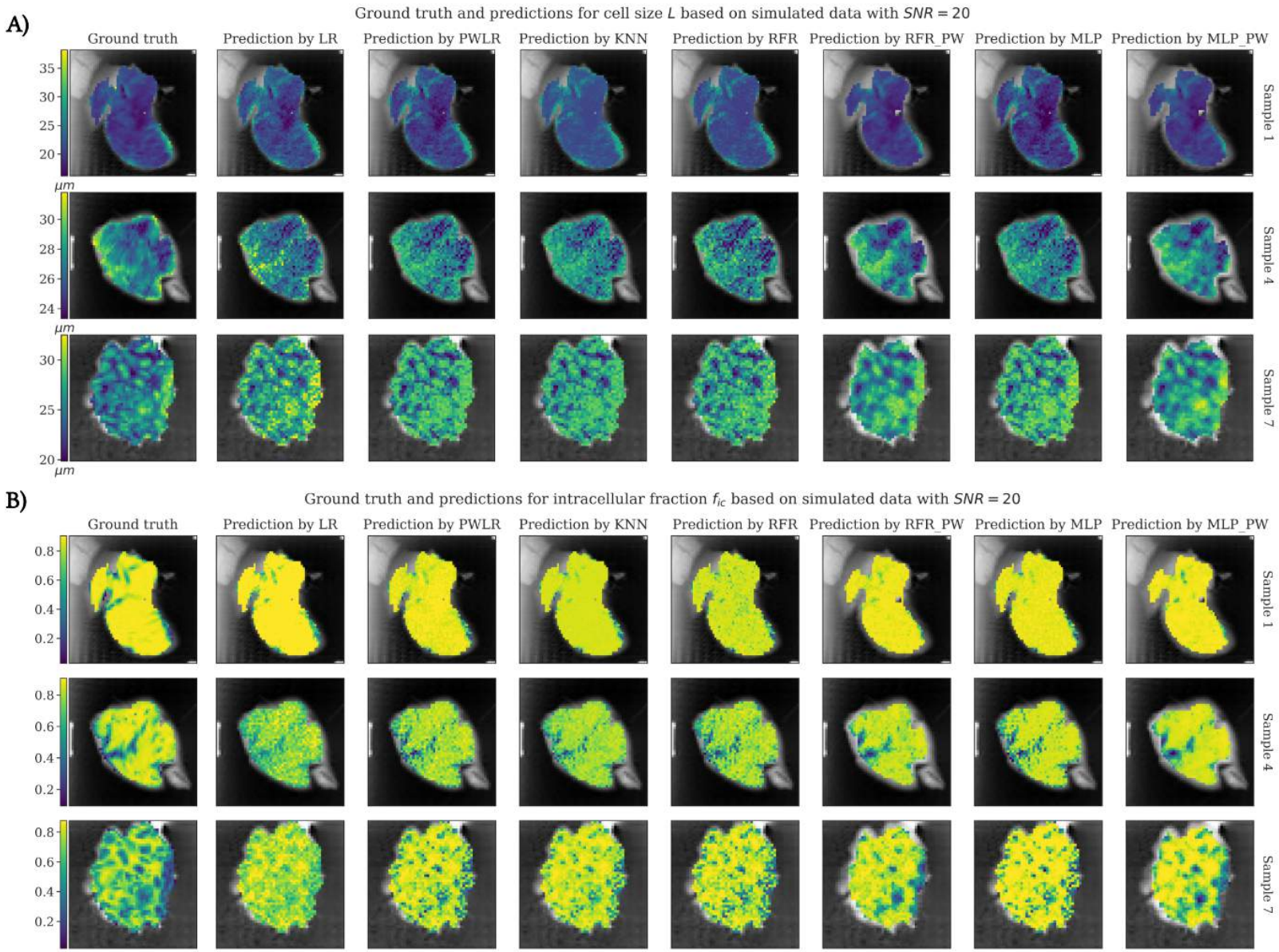


Figure B.1: The test predictions of (A) cell size L and (B) intracellular fraction f_{ic}, based on simulated data with $SNR = 20$, for from top to bottom WT, PDX_{inf1}, PDX_{nec}, and for from left to right predicted by Linear Regression (LR), Piecewise Linear Regression (PWLR), KNN (K-Nearest Neighbour), Random Forest Regression (RFR), Patch-wise RFR (PW-RFR), Multi-Layer Perceptron (MLP), and PW-MLP.

C

Results of the pre-clinical study

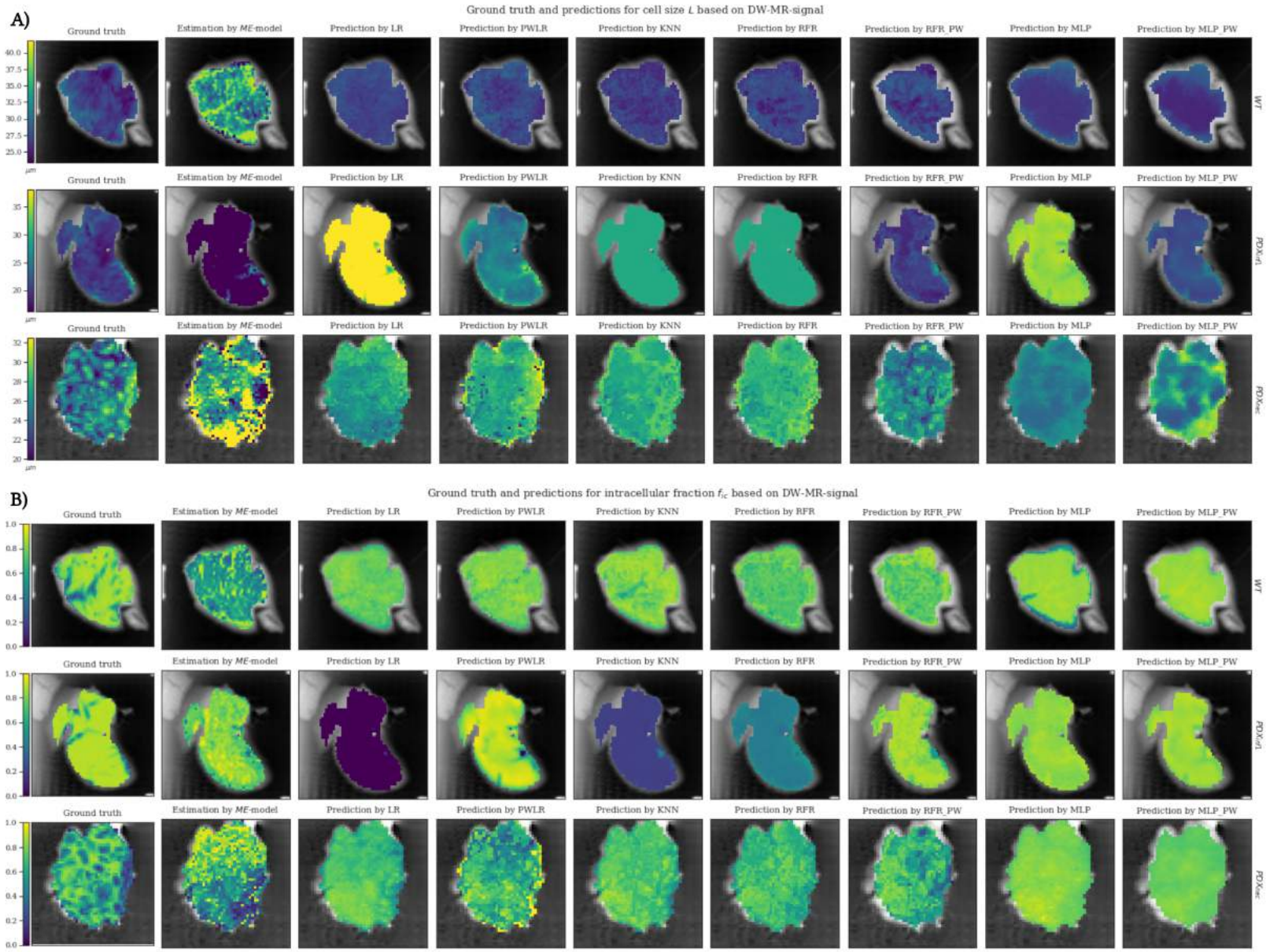


Figure C.1: The test predictions of (A) cell size L and (B) intracellular fraction f_{ic}, based on feature type *(i)* the DW-MR-signal, for from top to bottom WT, PDX_{infl}, PDX_{nec}, and for from left to right the ground truth, the estimation by the ME-model, and the prediction by Linear Regression (LR), Piecewise Linear Regression (PWLR), KNN (K-Nearest Neighbour), Random Forest Regression (RFR), Patch-wise RFR (PW-RFR), Multi-Layer Perceptron (MLP), and PW-MLP.

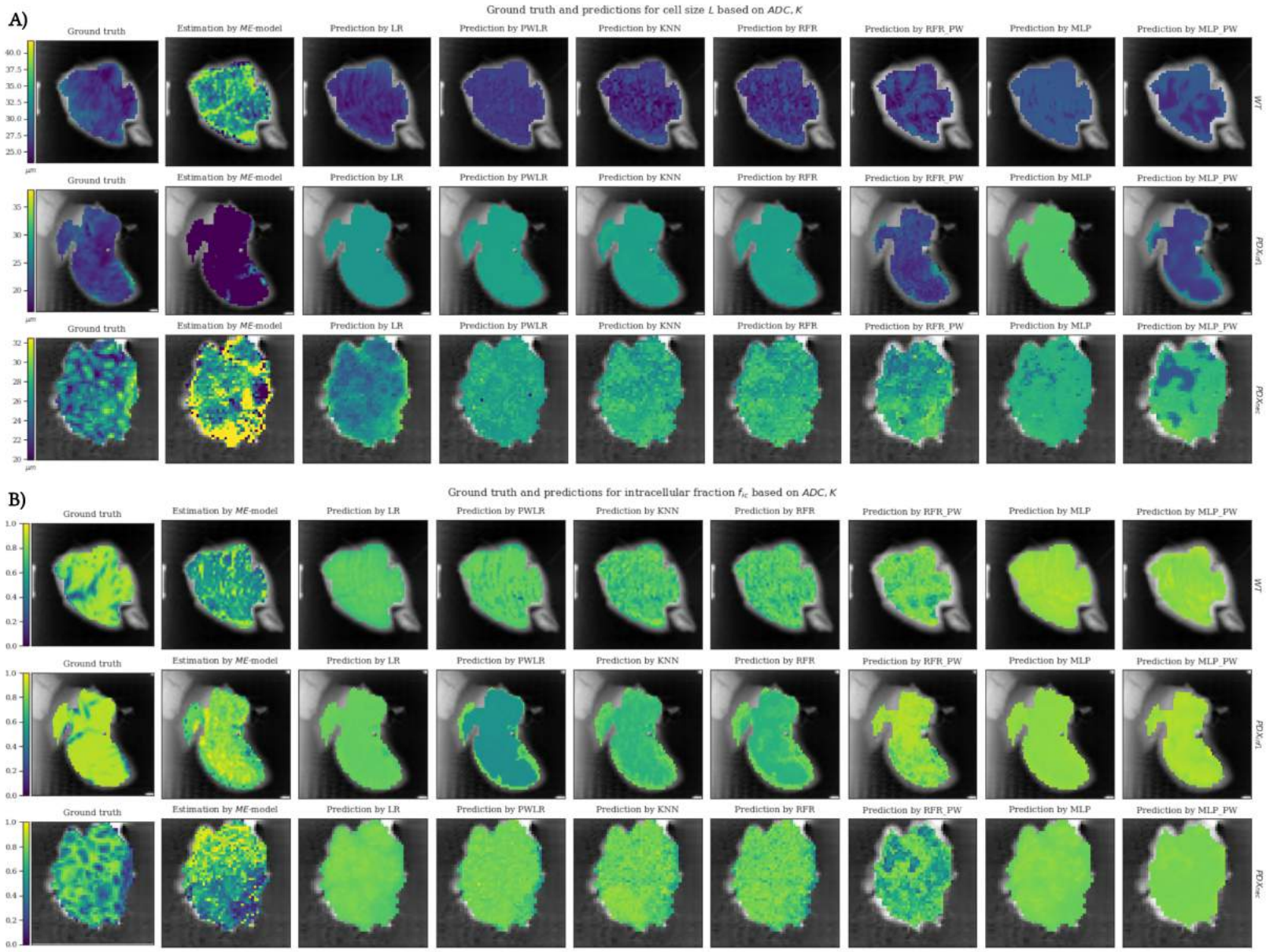


Figure C.2: The test predictions of (A) cell size L and (B) intracellular fraction f_{ic}, based on feature type *(ii)* $ADC - K$, for from top to bottom WT, PDX_{inf1}, PDX_{nec}, and for from left to right the ground truth, the estimation by the ME-model, and the prediction by Linear Regression (LR), Piecewise Linear Regression (PWLR), KNN (K-Nearest Neighbour), Random Forest Regression (RFR), Patch-wise RFR (PW-RFR), Multi-Layer Perceptron (MLP), and PW-MLP.

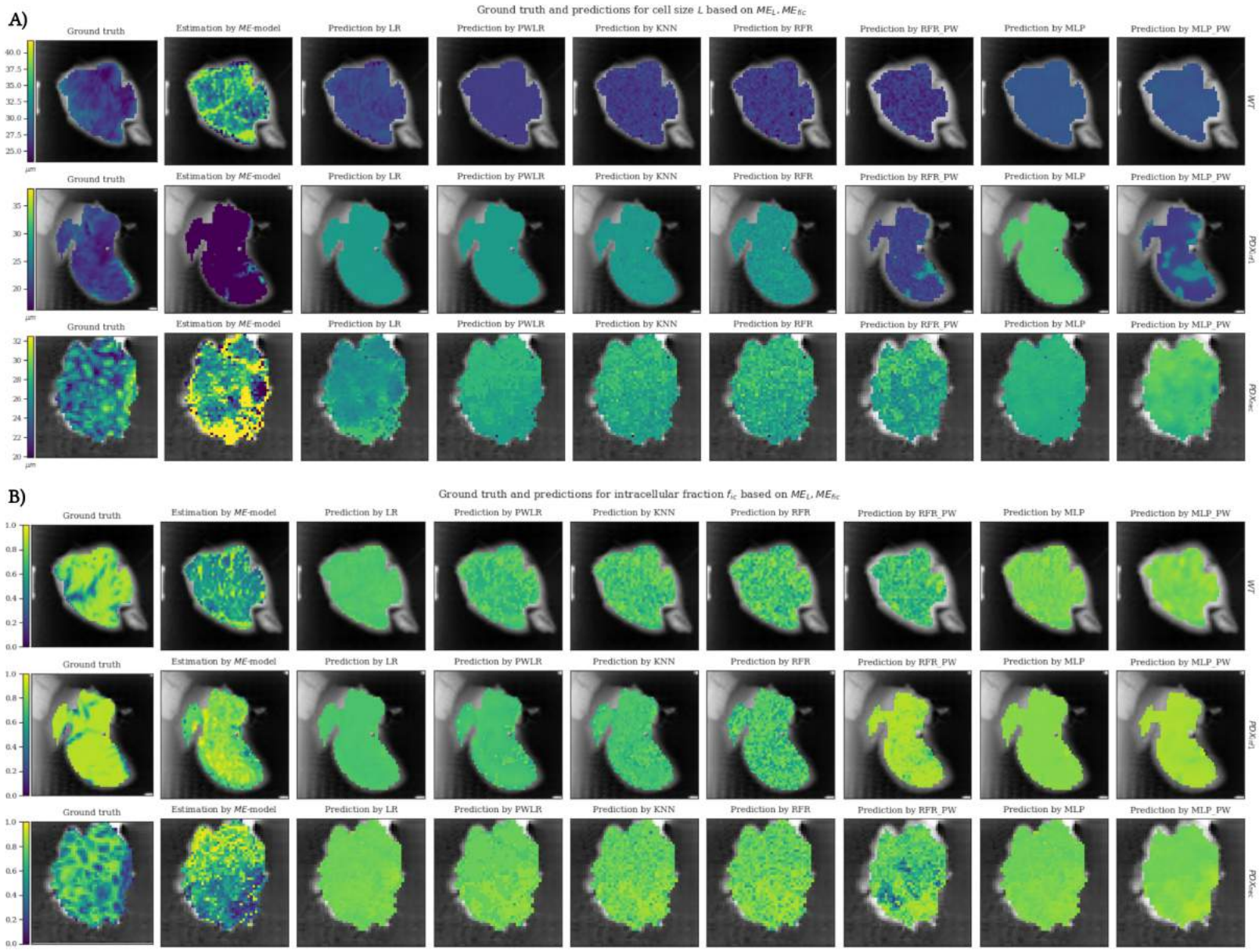


Figure C.3: The test predictions of (A) cell size L and (B) intracellular fraction f_{ic}, based on feature type *(iii)* the ME-estimations, for from top to bottom WT, PDX_{inf1}, PDX_{nec}, and for from left to right the ground truth, the estimation by the ME-model, and the prediction by Linear Regression (LR), Piecewise Linear Regression (PWLR), KNN (K-Nearest Neighbour), Random Forest Regression (RFR), Patch-wise RFR (PW-RFR), Multi-Layer Perceptron (MLP), and PW-MLP.