Learning microstructure parameters from diffusion-weighted MRI using random forests

Gemma Nedjati-Gilani[1], Matt G Hall[1], Claudia Angela M Wheeler-Kingshott[2], and Daniel C Alexander[1]
[1]*Computer Science & Centre for Medical Image Computing, University College London, London, London, United Kingdom,* [2]*Department of Neuroinflammation, Institute of Neurology, UCL, London, United Kingdom*

Target audience: Researchers interested in investigating and estimating indices of permeability; researchers interested in using machine learning techniques for medical image analysis.

Purpose Numerous white matter pathologies of the human central nervous system (CNS), such as multiple sclerosis, spinal cord injury and leukodystrophies, are characterised by myelin damage. As the breakdown of the myelin sheath is hypothesised to lead to an increase in axonal permeability, there is widespread interest in developing imaging biomarkers based on permeability in order to improve diagnosis and prognosis of these conditions. Diffusion tensor imaging (DTI) indices such as the radial diffusivity (RD) have been shown to correlate with myelin damage[1]; however other microstructure features also influence RD, thus reducing its specificity. The Kärger model[2] (KM) is a mathematical model that aims to directly relate the intracellular water exchange time to diffusion-weighted (DW) MR signals; however it relies on the assumption that the two pools of exchanging water are well mixed, which is not the case in white matter tissue where the intra and extra-cellular spaces are spatially separate. More recently, apparent exchange rate (AXR) imaging[3] has been introduced as an alternative to the KM. However, it still relies on strong assumptions about the compartmentation of water into a 'fast' and 'slow' pool. Given the inherent difficulties involved in deriving analytic models of permeability, in this study we construct a computational model using Monte Carlo (MC) simulations and machine learning. We use a regression forest to learn the mapping from simulations and obtain an efficient and accurate model for microstructure imaging that accounts for permeability. Previous work by Nilsson et al[4] generated libraries of microstructure parameters and their corresponding DW MR signals from MC simulations, and used them to find the nearest-neighbour microstructure parameters that matched unseen signals; however nearest-neighbour matching typically has poor generalization to unseen input data. Here we extend this approach using random forest regression[5], which has much better generalization to unseen data, i.e. combinations of tissue parameter values not explored in the training set.

Methods Monte Carlo Simulations: We use MC simulations to generate DW MR signals from 2500 substrates using the diffusion simulator in Camino[6]. White matter is modelled as a collection of non-abutting, parallel cylinders with radii drawn from a gamma distribution (with mean μ_R and standard deviation σ_R), and each substrate is described by a unique combination of μ_R, σ_R, volume fraction f and permeability p. The microstructure parameters for each substrate are randomly selected in the ranges: $\mu_R \in [0.2,5]$ μm, $\sigma_R \in [\min(0.1,\mu_R/5), \mu_R/2]$ μm (this is to ensure that the axon radius distributions cover the range observed in histology[7]), $f \in [0.4,0.7]$, $p \in [1\text{x}10^{-3},1\text{x}10^{-6}]$ ms^{-1}. This results in average intracellular water residence times of 8-200 ms, which are typical for biological tissue[8]. Diffusivity d is kept constant at $2\text{x}10^{-9}$ m^{2}s^{-1}, as we do not try to learn the relationship between d and the DW MR signals here. We simulate data using a rich acquisition protocol using all possible combinations of: G={0.02, 0.04, 0.06, 0.08, 0.1, 0.12, 0.14, 0.16, 0.18, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1} mTm^{-1}, Δ={5, 10, 20, 50, 100, 200, 500} ms, δ={2, 5} ms. We use DW directions both perpendicular and parallel to the cylinders. Each substrate contains 100,000 cylinders and all simulations are performed using 100,000 walkers and 1000 time steps. We generate two sets of signals: noise-free and SNR=20 (for the non-DW signals). Random Forest Regression: We use the scikit-learn machine learning toolbox[9] to train a random forest regressor, following the approach of Breiman[5], on 2000 of the 2500 generated signal vectors. The remaining 500 signal vectors are used for testing. The regressor is trained separately for the noise free and noisy data. The forest contains 500 estimators (trees). The maximum depth of each tree is automatically determined during training.

Results Figure 1 shows scatter plots of the volume fraction f, axon radius index[10] α (which reflects both the mean and the spread of the axon radius distribution) and permeability p against the values predicted by the random forest regressor for a) noise-free and b) SNR=20 data. Even though the data in a) is noise-free, due to the inherent statistical nature of the modelling, we do not observe an exact one-to-one correspondence between the predicted and ground truth microstructure parameters. The correlations for all parameters are strong, as indicated by the correlation coefficients. There is one clear outlier, which is highly visible in the scatter plots for f and α. The high permeability and small axon radius for this point result in a very short intracellular residence time of 8 ms, resulting in effectively free water diffusion over the diffusion times in all measurements in our protocol. Thus the random forest regressor is unable to estimate accurate values of f and α in this case. In b) the correlations are weaker for all parameters and we see that large values of f are often underestimated. However, correlation coefficients remain reasonably strong.

Discussion The results of this study demonstrate that microstructure parameters, including membrane permeability, can be learnt from simulated DW MR signals using random forest regression, even when the data is noisy. The strong correlation between predicted and ground truth permeabilities, even when using noisy data, is promising given the inherent difficulties in estimating this important microstructural parameter using existing mathematical models. In future, we intend to extend this work further. Feature selection will allow us to identify the signals that are most important for prediction, thus helping to reduce the protocol used in this experiment. We also intend to compare the approach with existing state of the art models and other regression techniques. The model we have learnt here is specifically for randomly packed, parallel, non-abutting cylinders; however the approach can be easily extended to other tissue configurations, including those with more than two compartments.

References [1]Song, NeuroImage 2005 [2]Karger, Adv in MR 1989 [3]Nilsson, MRM 2012 [4]Nilsson, ISMRM 2009 [5]Breiman, Ann Stats 1998 [6]Hall TMI 2009 [7]Aboitiz, Brain Res 1992 [8]Stein, 1967 [9]Pedregosa, JMLR 12 2011 [10]Alexander, NeuroImage 2010

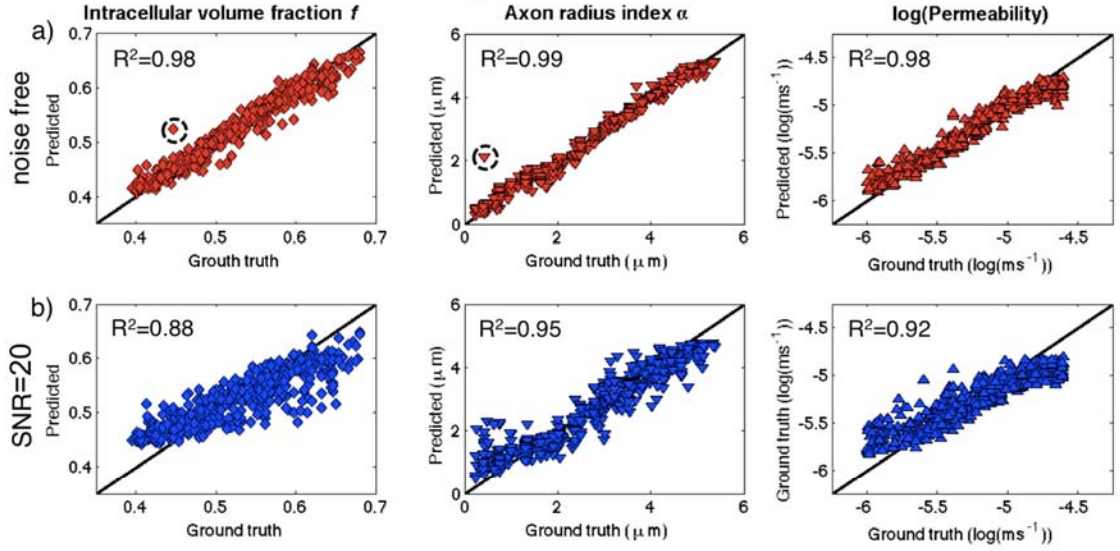

Figure 1: Scatter plots of ground truth versus predicted microstructure parameters for a) noise-free data and b) SNR=20 data.