NeuroTEST - How do image acquisition shifts affect SOTA MS lesion segmentation models?

Posselt Christiane, Yigitsoy Mehmet[1], Parida Abhijeet, Avci Mehmet Yigit, Stürzer Christina, Patrick Schünke, Christoph Kolbitsch, Tobias Schäffter, Remmele Stefanie [1]deepc GmbH, [2]Physikalisch Technische Bundesanstalt, Berlin

Purpose

This work investigates how the performance metric of an AI model for segmentation of multiple sclerosis (MS) lesions in T2-weighted FLAIR MR images changes in dependence of imaging scan parameters.

Introduction

- Several machine learning products have passed the approval process, e.g. to support radiologists with the diagnosis of med. images [1] MRI protocols are hardly standardized (see fig. 1), training images can differ a lot from images in application
- AI models cannot be tested against all possible image representations due to the lack of real data
- However, acquisition shifts can be described by the MR physics.
- Idea: simulation of possible image representations delivered by range of MR protocols → test data sets (Test case: T2w-FLAIR brain scans)
- model performance can be measured as a function of changing sequence parameters → prediction of AI performance in dependence of protocol

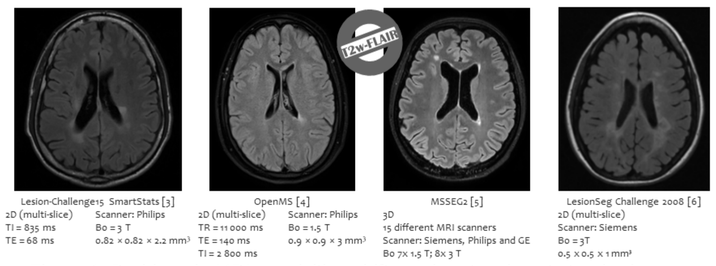

Figure 1. Problem: in MRI acquisition shifts and protocols not standards

Method

Research questions

Image simulation

- How realistic can domain shifts be simulated to generate test data for AI models?

Model analysis

- How do SOTA models behave in the presence of acquisition shifts?

Concept

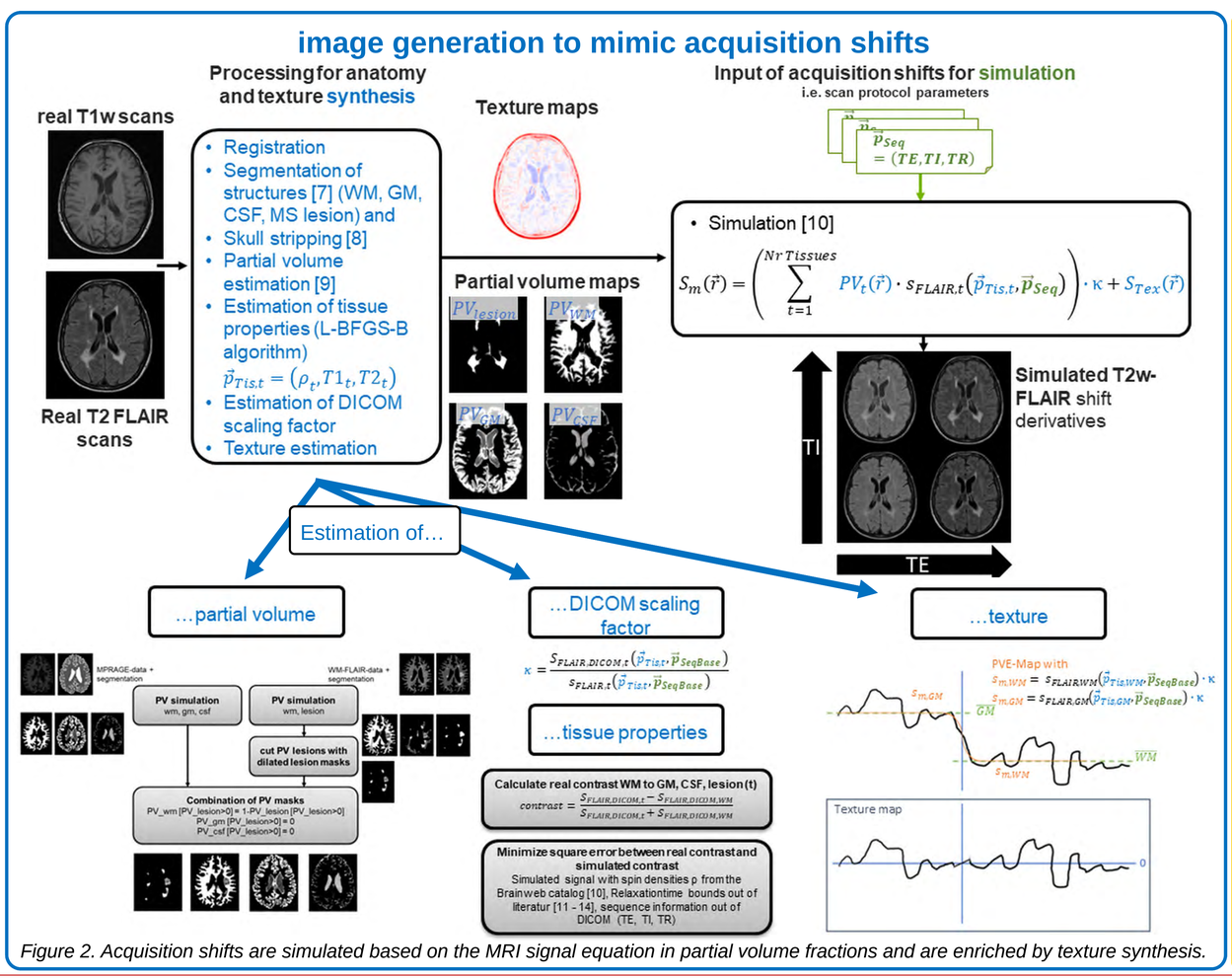

Figure 2. Acquisition shifts are simulated based on the MRI signal equation in partial volume fractions and are enriched by texture synthesis.

Modell stresstests with artificial data

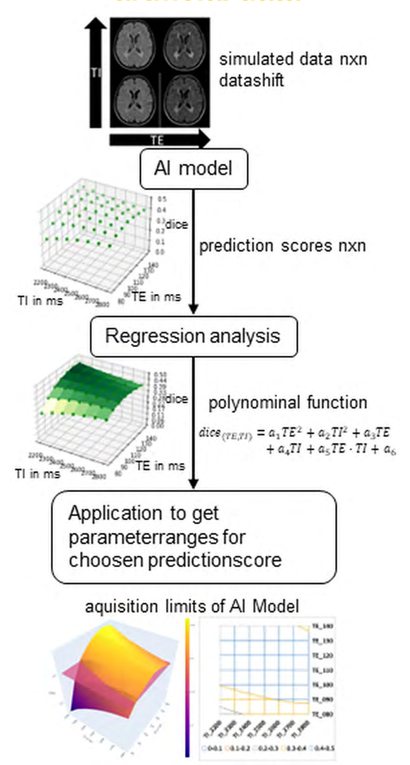

Figure 3. The AI model undergoes testing using generated images that have varying acquisition shifts. Through regression analysis, a function will be developed to provide the user with an assessment of the model's limitations.

Validation

Simulation Accuracy by:

Reference measurement

- comparison of simulation and real measurement of the same test person
- real measurements: 9 healthy patients; 3 T; variation of TE and TI

Stresstest *model* dependency

- Different SOTA models: nnUNET, SegResNet, UNETR, Vnet [15-18]
- Modelling performance (F1) in dependence of key sequence parameters (TI, TE, TR = 9000 ms)

Results

Reference measurement

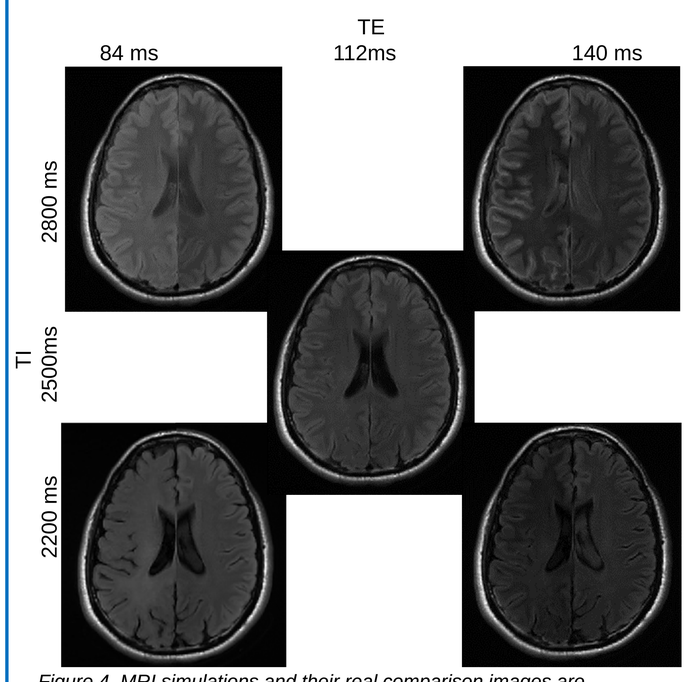

Figure 4. MRI simulations and their real comparison images are contrasted here with selected acquisition parameters. (left fake, right real)

Table 1. Comparison of the mean signals of WM, GM, CSF and skull of simulation and reference MRI with relative error in %

TE/ms	TI/ms	WM	GM	CSF	Skull
150	2900	18% ±6%	7% ±3%	75% ±30%	13% ±4%
150	2200	19% ±6%	9% ±7%	36% ±9%	25% ±2%
112	2500	0% ±0%	0% ±0%	0% ±0%	0% ±0%
84	2900	13% ±6%	8% ±6%	22% ±13%	21% ±2%
84	2200	12% ±6%	8% ±5%	58% ±10%	12% ±1%

Table 2. Exemplary relaxation parameters from a patient (optimised and real)

Relaxation times / ms	T1 wm	T2 wm	T1 gm	T2 gm	T1 csf	T2 csf
optimized on TE=112 ms, TI=2500 ms	1006	83	1773	135	4380	759
Reference measurements	1000	101	1562	117	3991	895

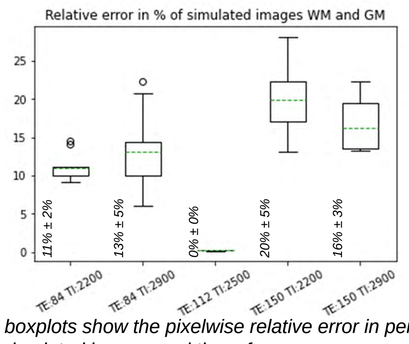

Figure 5. the boxplots show the pixelwise relative error in percent between the simulated images and the reference measurements in WM and GM of several datashifts of simulated images

Stresstest model dependency

Table 3. Coefficient of determination of the model fit (2nd order polynomial) to the measured dice scores.

	nnUNET	SegResNet	UNETR	VNet
Coefficient of determination R^2	0.985	0.983	0.980	0.982

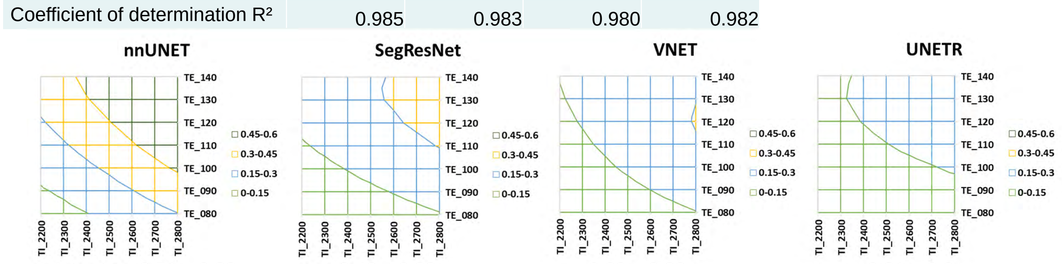

Figure 6. The elevation maps show the behaviour of the AI networks in dependence of the data shifts. The graphs show the dice trend as a function of the acquisitions parameters TE and TI. The real data dice comes from the baseline data with TE = 140 ms, TI = 2800 ms, TR = 11000 ms

Discussion

Image Simulation

- GM's simulation is the most successful with the smallest error (see Table 1 , Figure 4)
- CSF shows particularly high relative errors in the simulation (see Table 1, Figure 4) due to low signal in the baseline scan (“Fluid attenuated”!)
- It remains unclear whether average signal deviation (Fig. 1) reflect simulation errors or incorrect relaxation time estimation (Table 2). ,
- Pixelwise errors are also influenced by imperfect PV estimation (see Figure 5).
- The simulation method is also applicable to other sequences using the appropriate signal equations.
- A great advantage to GANs etc: arbitrary acquisition domains can be simulated not only the one of a particular target training domain.

Limitations:

- Simulation of only white matter, gray matter and CSF. The skull is stored in the texture map and is thus added back to the image -> expansion of the PV analysis needed
- method requires an additional T1w scan and the sequence parameters of the baseline T2 Flair scan -> T1w scans are recommended by all quality guidelines for neuroimaging
- exact validation not possible due to the lack of an accurate relaxometry reference (see Table 2) -> however, exact estimation not crucial for final simulation

Stresstest

- All SOTA models show dependence on acquisition shifts.
- Similar behavior observed among the models regarding range shifts (Figure 6)
- Higher contrast (higher TE and TI) results in better scores
- Previous work demonstrated relevant performance drops with shifts of TE and TI in training data
- Improvements made in the simulation, including the use of PV maps for tissue transitions and relaxation parameters based on original images
- Table 1 shows that all four SOTA models can be described with at least 98% of the fitted 2nd degree polynomial function of regression analysis
- The model appears suitable for predicting performance in changing scan situations
- Manufacturers could use this solution to recommend MRI image acquisition parameters to customers or to predict a performance drop/increase in a new image domain

Future work

- Uncertainty analysis: In order to take the AI model’s point of view when analyzing images.
- Stresstest training domain dependency.

Literature

[1] McCrindle, B., et al., *Radiology: Artificial Intelligence* 3(6), e210031, 2021
[2] P. Berlit, Hrsg., *Klinische Neurologie: Mit 363 Tabellen*, Berlin and Heidelberg: Springer, 2011
[3] ACR, List of FDA cleared AI medical products, https://aicentral.acrdsi.org/, abgerufen am 24.02.2022
[4] …
[5] …
[6] …

Nutzen Sie die Gelegenheit und probieren Sie doch mal Citavi oder Zotero aus! Achten Sie bei Webquellen auf das Datum. Benutzen Sie für Journal, Buch, Link die **Zitierstile**:

- NurErstautorname, V., et al, *Journal*, Volume(Heft), Seiten/e-Nr., Jahr.)
- NurErstautorname, V., et al, *Buchtitel*, Verlag, Jahr
- NurErstautor oder Hrsg, ggfs. Titel, URL, abgerufen am
- **WICHTIG**: Die Verwendung von OpenSource Tools und Datensätzen erfordert normalerweise die Zitation einer bestimmten Quelle, die Sie auf der jeweiligen Webseite finden!!!!!!!

Acknowlegement

This project was funded by the Federal Ministry of Economy and Technology (Project ZIM KK5050201LB0) and takes place in cooperation with the company deepc GmbH[1], the PTB Berlin and the LAKUMED hospital in Landshut.