Title of the Paper

Michael Mueller, Jan Riedo, Michael Rebsamen

Abstract—Bla

Index Terms—**MRI, Segmentation, Machine Learning, DF, kNN, SVM**

I. INTRODUCTION

Segmentation of brain tissues from magnetic resonance images (MRI) has many clinical applications. Clinicians gain useful information from a separation of tissue into its three main anatomical types: white matter, grey matter, and ventricles. However, manual segmentation of MRI is a labour-intensive task requiring expert skills. Fully automatic approaches for brain tissue segmentation are therefore a topic of active research. A good algorithm classifies the tissue types with high accuracy across a variety of images from different patients. Such a classification is a typical task for machine learning. These algorithms tend to perform well given enough training data during the learning phase. The availability of ground-truth data in sufficient quantity and quality for supervised learning is a particular challenge when working with medical images due to privacy concerns and the costs for manual segmentation. Optimization of the learning phase with a limited number of training data is therefore required.

FIXME: kNN is a popular classification method for MR data and has successfully been applied in MR brain segmentation [1]–[3]

FIXME: Base paper on df [4].

II. METHODS

A. Dataset

All experiments were conducted on a subset of 100 unrelated subjects from a dataset provided by the *Human Connectome Project* [5]. From each individual, a total of eight 3-tesla head MRI are available: T1 and T2-weighted image volumes not skull-stripped (but defaced for anonymization) and skull-stripped with a bias field correction, and both modalities once in native T1 space and once in MNI-atlas space [6].

Ground-truth labels are automatically generated using *FreeSurf*, assigning each voxel either to background, white matter, grey matter, or ventricles. The dataset was split in a training set with 70 images and a test set with 30 images.

B. Pipeline

TODO: Describe whole pipeline (registration, preprocessing, feature extraction, ML classification, postprocessing, evaluation FIXME: During feature extraction, a random mask is applied in order to randomly select a fraction

Biomedical Engineering, University of Bern
Authors e-Mail: michael.mueller@students.unibe.ch, jan.riedo@students.unibe.ch, michael.rebsamen@students.unibe.ch

of the voxels available. The mask is adjustable individually for background, white matter, grey matter, and ventricles. It is crucial to optimize those parameters in order to obtain the best result.

C. Training

TODO: Describe training of machine learning algorithms TODO: SVM gridsearch for hyperparameter tuning?

D. Performance Evaluation

TODO: Describe metric (dice score)

E. Infrastructure

TODO: Describe UBELIX, libraries

III. RESULTS

One major task to handle was the low value for the ventricles. Dice values above 0.5 were hard to achieve. One way to improve the dice for ventricles was to optimize the random mask with respect to the fraction of ventricle voxels taken into account. The effects of the random mask on the ventricle dice can be seen in Fig. (1). Best results were achieved with a fraction of 0.004 ventricles, approximately the same fraction as for white matter and grey matter. All following results are based on this optimized mask.
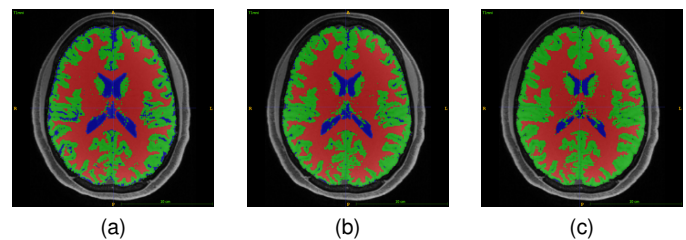

(a) (b) (c)

Fig. 1. Optimization of the random mask parameter for ventricles. Fraction of ventricle voxels taken into account f_v and dice value for this certain parameter d_v are: (f_v / d_v) (a) 0.4 / 0.22, (b) 0.04 / 0.44, and (c) 0.004 / 0.62.

TODO: JR: subplot with 3 result images (good dice bad result, good dice good result, bad ground truth

TODO: JR: DF hyperparameter optimization, 3DPlot The decision forest algorithm was enhanced with normalized features, a higher number of ventricle voxels in the training set and the optimization of the hyperparameters (see Fig. III). With this settings, the max dice coefficient was lifted from 0.703 to 0.754. This result was achieved with 80 trees and 3000 max nodes.

TODO: JR: kNN optimization

Statistical distribution of the dice coefficients can be seen in Fig. III. DF and SVM achieve a similar mean dice score but SVM has a lower variance for the ventricles.

a

TABLE I
PERFORMANCE COMPARISON OF ML ALGORITHMS

Features	Size Dataset	DF	GMM	kNN	SGD	SVM
All (f1-f7)	3	-	-	0.70/0.57/0.48	-	-
	12	0.84/0.80/0.52	0.00/0.78/0.00	0.75/0.66/0.67	-	0.83/0.81/0.58
	70	0.85/0.80/0.60	-	0.79/0.76/0.73	0.82/0.80/0.33	0.84/0.82/0.61
Coordinates only (f1-f3)	3	-	-	0.70/0.55/0.41	-	-
	12	-	-	0.74/0.63/0.56	-	-
	70	-	-	0.77/0.71/0.62	-	-
All non-coordinates (f4-f7)	3	-	-	0.85/0.80/0.45	-	-
	12	-	-	0.85/0.81/0.45	-	-
	70	-	-	0.85/0.81/0.54	-	-

Overview of achieved accuracy for the different algorithms. Mean dice scores for white matter/grey matter/ventricles.
f1-f3: Coordinate features, f4: T1 intensity, f5: T1 gradient, f6: T2 intensity, f7: T2 gradient.

TABLE II
RUNTIME

Features	Size Dataset	DF	GMM	kNN	SGD	SVM
All (f1-f7)	3	-	-	14.4/6564.6	-	-
	12	-	-	39.6/7548.7	-	-
	70	-	-	223.1/8724.6	-	-
Coordinates only (f1-f3)	3	-	-	10.4/4391.5	-	-
	12	-	-	34.7/5449.3	-	-
	70	-	-	196.4/6112.8	-	-
All non-coordinates (f4-f7)	3	-	-	10.1/10084.7	-	-
	12	-	-	34.6/18768.6	-	-
	70	-	-	194.2/16555.7	-	-

FIXME: Overview of the computation time in seconds for all algorithms (training time/testing time). Computation time includes pre- and post-processing.
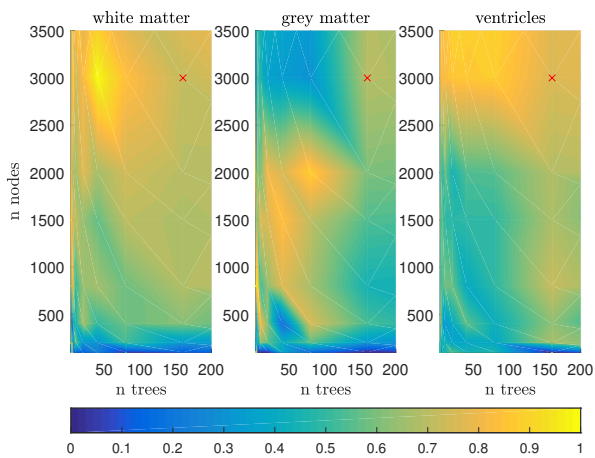

Fig. 2. DF plot of grid search for white matter, grey matter and ventricles. The red cross marks the chosen hyperparameters number of trees = 160 and maximum nodes per tree = 3000. Color does not represent dice, the data is stretched individually for all three plots.

Fig. 3. Scatter plot of the features with correlation coefficient

V. CONCLUSION

ACKNOWLEDGMENT

IV. DISCUSSION

TODO: challenge with quality of ground truth

TODO: feature importance

REFERENCES

[1] P. Anbeek, K. L. Vincken, M. J. van Osch, R. H. Bisschops, and J. van der Grond, "Probabilistic segmentation of white matter lesions in MR imaging," *NeuroImage*, vol. 21, no. 3, pp. 1037–1044, mar 2004.
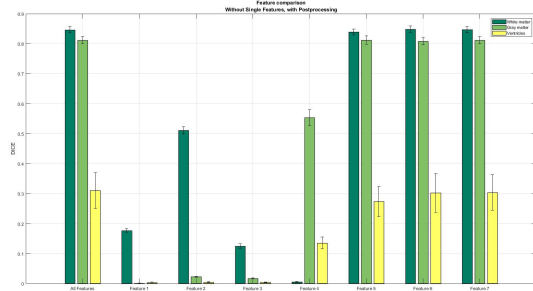
Fig. 4. Feature evaluation with Decision Forest by removing a single feature and with preprocessing
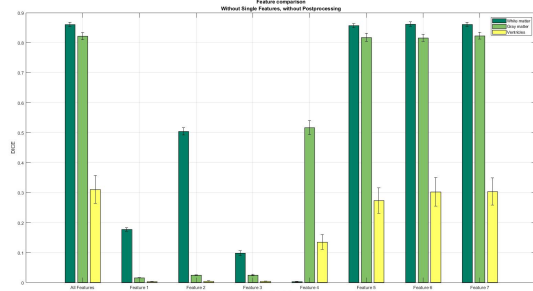

Fig. 5. Feature evaluation with Decision Forest by removing a single feature and without preprocessing
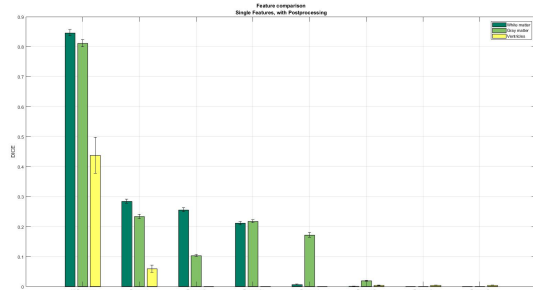

Fig. 6. Feature evaluation with Decision Forest by using a single feature and with preprocessing
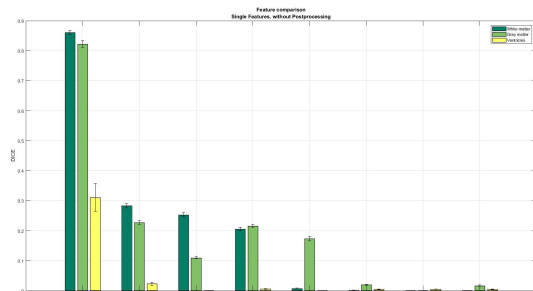

Fig. 7. Feature evaluation with Decision Forest by using a single feature and without preprocessing
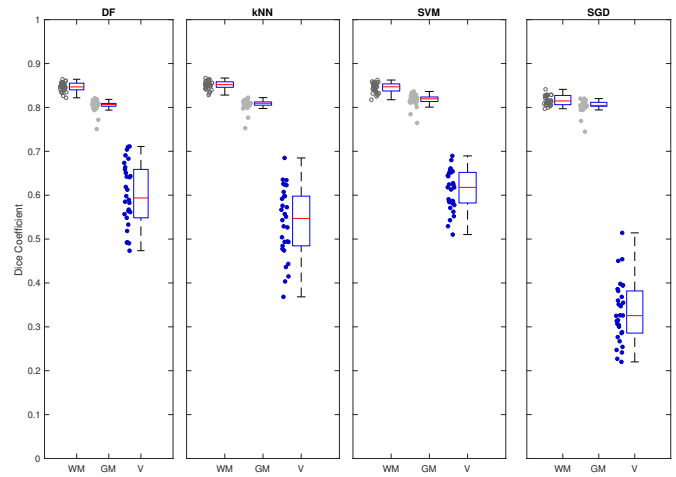

Fig. 8. Distribution of dice coefficients with optimal hyper-parameters for each algorithm on the full training set of 70 images.
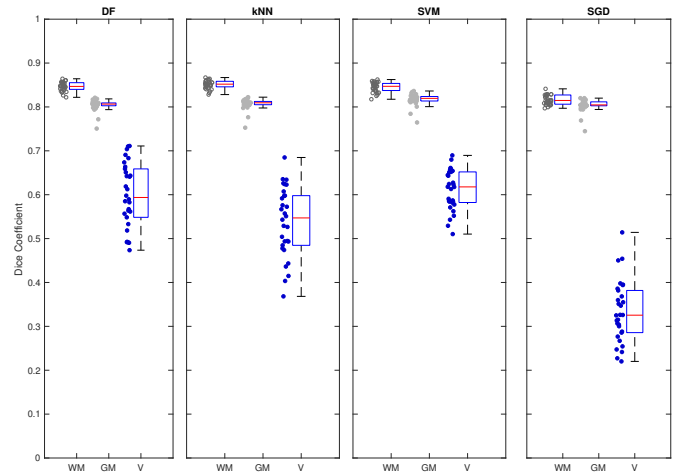

Fig. 9. Distribution of dice coefficients with optimal hyper-parameters for each algorithm on the full training set of 70 images.

[2] C. A. Cocosco, A. P. Zijdenbos, and A. C. Evans, "A fully automatic and robust brain MRI tissue classification method," *Medical Image Analysis*, vol. 7, no. 4, pp. 513–527, dec 2003.

[3] S. Warfield, M. Kaus, F. A. Jolesz, and R. Kikinis, "Adaptive, template moderated, spatially varying statistical classification," *Medical Image Analysis*, vol. 4, no. 1, pp. 43–55, mar 2000.

[4] L. Breiman, *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[5] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, K. Ugurbil, W.-M. H. Consortium *et al.*, "The wu-minn human connectome project: an overview," *Neuroimage*, vol. 80, pp. 62–79, 2013.

[6] J. Mazziotta, A. Toga, A. Evans, P. Fox, J. Lancaster, K. Zilles, R. Woods, T. Paus, G. Simpson, B. Pike *et al.*, "A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (icbm)," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 356, no. 1412, pp. 1293–1322, 2001.