

Spatial Decision Forests for MS Lesion Segmentation in Multi-Channel MR Images

Ezequiel Geremia^{1,3}, Bjoern H. Menze^{1,2}, Olivier Clatz¹, Ender Konukoglu³, Antonio Criminisi³, and Nicholas Ayache¹

¹ Asclepios Research Project, INRIA Sophia-Antipolis, France

² Computer Science and Artificial Intelligence Laboratory, MIT, USA

³ Machine Learning and Perception Group, Microsoft Research Cambridge, UK

Abstract. A new algorithm is presented for the automatic segmentation of Multiple Sclerosis (MS) lesions in 3D MR images. It builds on the discriminative random decision forest framework to provide a voxel-wise probabilistic classification of the volume. Our method uses multi-channel MR intensities (T1, T2, Flair), spatial prior and long-range comparisons with 3D regions to discriminate lesions. A symmetry feature is introduced accounting for the fact that some MS lesions tend to develop in an asymmetric way. Quantitative evaluation of the data is carried out on publicly available labeled cases from the MS Lesion Segmentation Challenge 2008 dataset and demonstrates improved results over the state of the art.

1 Introduction

Multiple Sclerosis (MS) is a chronic, inflammatory and demyelinating disease that primarily affects the white matter of the central nervous system. Automatic detection and segmentation of MS lesions can help diagnosis and patient follow-up. It offers an attractive alternative to manual segmentation which remains a time-consuming task and suffers from intra- and inter-expert variability. MS lesions, however, show high appearance variability which makes automatic segmentation a challenging task. Indeed, MS lesions lack common intensity and texture characteristics, their shapes are variable and their location within the white matter varies across patients.

A variety of methods have been proposed for the automatic segmentation of MS lesions. Generative methods were proposed consisting in a tissue classification by means of an expectation maximization (EM) algorithm. The EM algorithm can be modified to be robust against lesion affected regions, its outcome is then parsed in order to detect outliers which, in this case, coincide with MS lesions [1]. Another approach consists in adding to the EM a partial volume effect model between tissue classes and combining it with a Mahalanobis thresholding which highlights the lesions [2]. Morphological postprocessing on resulting regions of interest was shown to improve the classification performance [3]. In [4], a constrained Gaussian mixture model is proposed, with no spatial prior, to capture the tissue spatial layout. MS lesions are detected as outliers

and then grouped in an additional tissue class. Final delineation is performed using probability-based curve evolution. Multi-scale segmentation can be combined with discriminative classification to take into account regional properties [5]. Beyond the information introduced via the spatial prior atlases, these methods are limited in their ability to take advantage of long-range spatial context in the classification task.

To overcome this shortcoming, we propose the use of an ensemble of discriminative classifiers. It builds on the random decision forest framework which has multiple applications in bioinformatics [6], but more recently also in the image processing community [7]. Adding spatial and multi-channel features to this classifier proved effective in object recognition [8], brain tissue segmentation in MR images [9], myocardium delineation in 3D echocardiography [10] and organ localization in CT volumes [11]. Applying multi-channel and context-rich random forest classification to the MS lesion segmentation problem is novel, to our knowledge. The presented classifier also exploits a specific discriminative symmetry feature assuming that the healthy brain is approximately symmetric with respect to the mid-sagittal plane and that MS lesions tend to develop in asymmetric ways.

2 Materials and Methods

This section describes our adaptation of the random decision forests to the segmentation of MS lesions and illustrates the visual features employed.

2.1 Dataset

Our dataset contains 20 labeled cases which are publicly available from the MS Lesion Segmentation Challenge 2008 website [12]. For each case, three MR channels are made available T1-, T2-weighted and Flair.

After being sub-sampled and cropped, all the images have the same size, $159 \times 207 \times 79$ voxels, and the same resolution, $1 \times 1 \times 2 \text{ mm}^3$. RF acquisition field inhomogeneities are corrected [13] and inter-subject intensity variations are normalized [14]. The images are then aligned on the mid-sagittal plane [15]. Spatial prior is added by registering the MNI atlas [16] to the anatomical images, each voxel of the atlas providing the probability of belonging to the white matter (WM), the grey matter (GM) and the cerebro-spinal fluid (CSF) (cf. Fig. 1).

We will adhere to the following notation: the data consists of a collection of voxel samples $\mathbf{v} = (\mathbf{x}, \mathbf{C})$, each characterized by a position $\mathbf{x} = (x, y, z)$ and associated with a list of signal channels \mathbf{C} . Signal channels $\mathbf{C} = (\mathbf{I}, \mathbf{P})$ include multi-sequence MR images $\mathbf{I} = (I_{T1}, I_{T2}, I_{Flair})$ and spatial priors $\mathbf{P} = (P_{WM}, P_{GM}, P_{CSF})$. Anatomical images and spatial priors, although having different semantics, can be treated under the unified term “signal channel”. We account for noise in MR images by averaging values over a 3^3 voxels box centered on \mathbf{x} , such an average is noted $C_c(\mathbf{x})$, e.g. $C_c = I_{Flair}$ or P_{GM} .

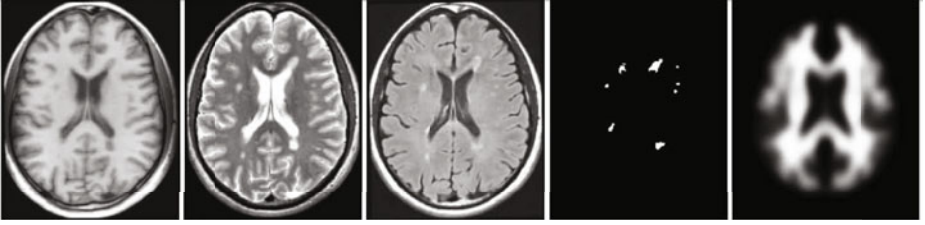


Fig. 1. Data. From left to right: preprocessed T1-weighted, T2-weighted and Flair MR images, the associated ground truth and the registered white matter atlas.

2.2 Context-Rich Decision Forest

Our detection and segmentation problem can be formalized as a binary classification of voxel samples into either background or lesions. This classification problem is addressed by a supervised method: discriminative random decision forest, an ensemble learner using decision trees as base learners. Decision trees are discriminative classifiers which are known to suffer from over-fitting. A random decision forest [17] achieves better generalization by growing an ensemble of many independent decision trees on a random subset of the training data and by randomizing the features made available to each node during training [18].

Forest training. The forest has T components with t indexing each tree. The training data consists in a set of labeled voxels $\mathcal{T} = \{\mathbf{v}_k, Y(\mathbf{v}_k)\}$ where the label $Y(\mathbf{v}_k)$ is given by an expert. When asked to classify a new image, the classifier aims to assign every voxel \mathbf{v} in the volume a label $y(\mathbf{v})$. In our case, $y(\mathbf{v}) \in \{0, 1\}$, 1 for lesion and 0 for background.

During training, all observations \mathbf{v}_k are pushed through each of the trees. Each internal node applies a binary test [8,9,10,11] as follows:

$$t^{\tau_{low}, \tau_{up}, \theta}(\mathbf{v}_k) = \begin{cases} true, & \text{if } \tau_{low} \leq \theta(\mathbf{v}_k) < \tau_{up} \\ false, & \text{otherwise} \end{cases}$$

where θ is a function identifying the visual feature extracted at position \mathbf{x}_k . There are several ways of defining θ , either as a local intensity-based average, local spatial prior or context-rich cue. These are investigated in more detail in the next section. The value of the extracted visual feature is thresholded by τ_{low} and τ_{up} . The voxel \mathbf{v}_k is then sent to one of the two child nodes based on the outcome of this test. Training the classifier means selecting the most discriminative binary test for each node by optimizing over $(\tau_{low}, \tau_{up}, \theta)$ in order to maximize the information gain on the input data partition [19], noted \mathcal{T}_p , defined as follows: $IG_{\tau_{low}, \tau_{up}, \theta}(\mathcal{T}_p) = H(\mathcal{T}_p) - H(\mathcal{T}_p | \{t^{\tau_{low}, \tau_{up}, \theta}(\mathbf{v}_k)\})$ where $\mathcal{T}_p \subset \mathcal{T}$, H stands for the entropy.

Only a randomly sampled subset Θ of the feature space is available for internal node optimization, while the threshold space is uniformly discretized. The optimal $(\tau_{low}^*, \tau_{up}^*, \theta^*)$ is selected by exhaustive search jointly over the feature and threshold space. Random sampling of the features leads to increased

inter-node and inter-tree variability which improves generalization. Nodes are grown to a maximum depth D . Another stopping criterion is to stop growing a node when too few training points reach it, i.e. when the information gain is below a minimal value IG_{min} .

As a result of the training process, each leaf node l of every tree t receives a partition \mathcal{T}_l of the training data. The following empirical posterior probability is then stored at the leaf $p_{l_t}(Y(\mathbf{v}) = b) = |\{(\mathbf{v}, Y(\mathbf{v})) \in \mathcal{T}_l | Y(\mathbf{v}) = b\}| / |\mathcal{T}_l|$ where $b \in \{0, 1\}$ denotes the background or lesion class, respectively.

Prediction. When applied to a new test data $\mathcal{T}_{test} = \{\mathbf{v}_k\}$, each voxel \mathbf{v}_k is propagated through all the trees by successive application of the relevant binary tests. When reaching the leaf node l_t in all trees $t \in [1..T]$, posteriors $p_{l_t}(Y(\mathbf{v}) = b)$ are gathered in order to compute the final posterior probability defined as follows: $p(y(\mathbf{v}) = b) = \frac{1}{T} \sum_{t=1}^T p_{l_t}(Y(\mathbf{v}) = b)$. This probability may be thresholded at a fixed value $T_{posterior}$ if a binary segmentation is required.

2.3 Visual Features

In this section, two kinds of visual features are computed: 1) local features: $\theta_c^{loc}(\mathbf{v}) = C_c(\mathbf{x})$ where c indexes an intensity or a prior channel; 2) context-rich features comparing the voxel of interest with distant regions. The first context-rich feature looks for relevant 3D boxes R_1 and R_2 to compare within an extended neighborhood: $\theta_{c_1, c_2, R_1, R_2}^{cont}(\mathbf{v}) = C_{c_1}(\mathbf{x}) - \frac{1}{vol(R_1 \cup R_2)} \sum_{\mathbf{x}' \in R_1 \cup R_2} C_{c_2}(\mathbf{x}')$ where c_1 and c_2 are two signal channels. The regions R_1 and R_2 are sampled randomly in a large neighborhood of the voxel \mathbf{v} (cf. Fig. 2). The sum over these regions is efficiently computed using integral volume processing [8]. The second context-rich feature compares the voxel of interest at \mathbf{x} with its symmetric counterpart with respect to the mid-sagittal plane, noted $S(\mathbf{x})$: $\theta_c^{sym}(\mathbf{v}) = C_c(\mathbf{x}) - C_c \circ S(\mathbf{x})$ where c is an intensity channel. Instead of comparing with the exact symmetric $S(\mathbf{x})$ of the voxel, we consider, respectively, its 6, 26 and 32 neighbors in a sphere

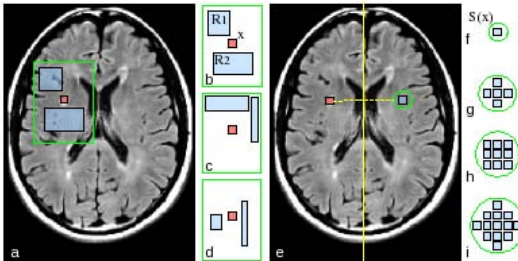


Fig. 2. 2D view of context-rich features. (a) A context-rich feature depicting two regions R_1 and R_2 with constant offset relatively to \mathbf{x} . (b-d) Three examples of randomly sampled features in an extended neighborhood. (e) The symmetric feature with respect to the mid-sagittal plane. (f) The hard symmetric constraint. (g-i) The soft symmetry feature considering neighboring voxels in a sphere of increasing radius. See text for details.

\mathcal{S} (cf. Fig. 2), centered on $S(\mathbf{x})$. We obtain a softer version of the symmetric feature which reads: $\theta_{c,S}^{sym}(\mathbf{v}) = \min_{\mathbf{x}' \in \mathcal{S}} \{C_c(\mathbf{x}) - C_c(\mathbf{x}')\}$.

3 Results

In our experiments, forest parameters are fixed to the following values; number of random regions $|\Theta| \simeq 950$, number of trees $T = 30$, tree depth $D = 20$, lower bound for the information gain $IG_{min} = 10^{-5}$, posterior threshold $T_{posterior} = 0.5$. These values were chosen based on prior parameter optimization on synthetic data and worked well for real data too.

For quantitative evaluation, the 20 available cases are classified and compared to one of the state of the art methods [3]. A three-fold cross-validation is carried out on this dataset: the forest is trained on $\frac{2}{3}$ of the cases and tested on the other $\frac{1}{3}$, this operation is repeated three times in order to collect test errors for each case. Note that the random forest is trained on the preprocessed data.

The binary classification is evaluated using two measures, true positive rate (TPR) and positive predictive value (PPV), both equal 1 for perfect segmentation. Formally, $TPR = \frac{TP}{TP+FN}$ and $PPV = \frac{TP}{TP+FP}$ where TP counts the number of true positive voxels in the classification compared to the ground truth, FP the false positives, FN the false negatives.

Random forest based segmentations are compared with an aligned and sub-sampled version of the ground truth (cf. Fig. 3), whereas segmentations from the winner algorithm were compared with original and sub-sampled segmentations for similar results. Our segmentation compares favorably to one of the state of the art algorithms (cf. Table 1) for both TPR and PPV .

The Grand Challenge 2008 website carried out a complementary and independent evaluation of the algorithm on their private dataset [12]. The results confirm a significant improvement over the winner algorithm of the challenge [3]. The presented spatial random forest achieves, in average, slightly larger true positive (TPR), which is beneficial, and comparable false positive (FPR) rates but significantly lower volume difference (VD), and surface distance (SD) values.

3.1 Discussion

Influence of preprocessing. Data normalization is critical. Indeed, features selected during training should be applied exactly in the same way to new data. For instance, context-rich features, θ^{cont} , are sensitive to rotation and thus require aligned images. Moreover, intensity based features require inter-image normalization to ensure consistency of threshold values in binary tests. This limitation is merely due to our supervised approach. On the contrary, image sub-sampling does not affect feature evaluation much as we are considering averages over rectangular regions.

Analysis of feature relevance. Decision trees are interesting because of ease of interpretability which is highly relevant in classification algorithms considering a large number of input features [6]. We exploit this to analyze the selected

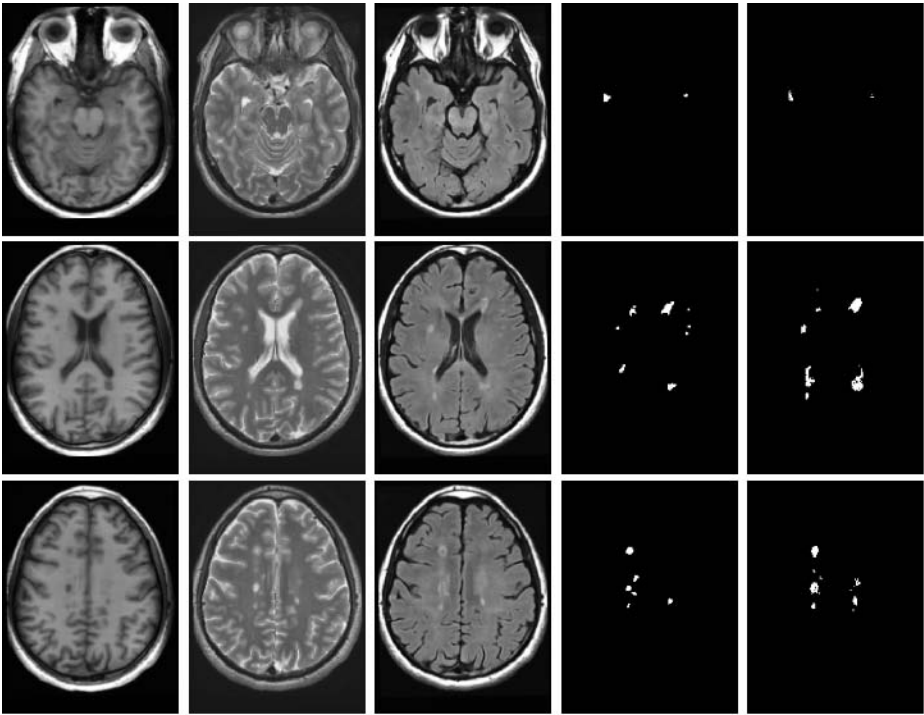


Fig. 3. Segmentation results on a multi-channel 3D MR image. Rows: Axial slices. Columns (from left to right): T1, T2, Flair MR images, ground truth and the output posterior after thresholding.

Table 1. Comparison of context-rich random forests with a state of the art method. In bold we indicate where we do better than the winner algorithm of the MS Segmentation Challenge 2008.

	Ch. winner [3]		Context-rich RF			Ch. winner [3]		Context-rich RF	
Patient	TPR	PPV	TPR	PPV	Patient	TPR	PPV	TPR	PPV
CHB01	0.22	0.41	0.49	0.64	UNC01	0.01	0.01	0.02	0.01
CHB02	0.18	0.29	0.44	0.63	UNC02	0.37	0.39	0.48	0.36
CHB03	0.17	0.21	0.22	0.57	UNC03	0.12	0.16	0.24	0.35
CHB04	0.12	0.55	0.31	0.78	UNC04	0.38	0.54	0.54	0.38
CHB05	0.22	0.42	0.40	0.52	UNC05	0.38	0.08	0.56	0.19
CHB06	0.13	0.46	0.32	0.52	UNC06	0.09	0.09	0.15	0.08
CHB07	0.13	0.39	0.40	0.54	UNC07	0.57	0.18	0.76	0.16
CHB08	0.13	0.55	0.46	0.65	UNC08	0.27	0.20	0.52	0.32
CHB09	0.03	0.18	0.23	0.28	UNC09	0.16	0.43	0.67	0.36
CHB10	0.05	0.18	0.23	0.39	UNC10	0.22	0.28	0.53	0.34

features and understand what are the most discriminative channels for MS lesion segmentation. For this analysis, we consider one of the random forest classifiers which generated the results in Table 1. For every tree in the forest, the root node always applies a test on the Flair sequence (θ_{Flair}^{loc}). It means that out of all available features, containing local, context-rich and symmetry multi-channel features, θ_{Flair}^{loc} was found to be the most discriminative. This automated guess coincides with the first step in [3]. At the second level of the tree, a context-rich feature on prior information ($\theta_{WM,GM}^{cont}$) appears to be the most discriminative over all trees in the forest. The associated test discards all voxels which do not belong to the white matter. Again, our algorithm automatically reproduced the second step in [3]. In deeper levels of the tree, local, context-rich and symmetry features adjust the segmentation by combining spatial and multi-channel information. Contribution of each feature to the forest can be quantified by counting the nodes in which they were selected. This indicates a feature discrimination power for the task of MS lesion classification. Local features were selected in 24% of the nodes, context-rich features were selected in 71% of the nodes whereas symmetry features were selected in only 5%. Successive decisions based on local features may learn a non-parametric multi-channel appearance model with spatial prior. Context-rich features exhibit high variability (900 of them are randomly sampled at every node). This variability combined with their ability to highlight regions which differ from their neighborhood explains the high selection frequency. In addition, this kind of features may learn a spatial layout for lesion patterns in peri-ventricular regions (cf. second row in Fig. 3). Symmetry features are under-represented in the forest and thus prove to be the least discriminative ones. Nevertheless, they appear in top levels of the tree (up to third level) which indicates that, they provide an alternative to local and context-rich features when these two fail.

4 Conclusion

We introduce a new algorithm for the segmentation of MS lesions in multi-channel MR images. We present three kinds of 3D features based on multi-channel intensity, prior and context-rich information. Those features are part of a spatial random decision forest classifier which demonstrates improved results on one of the state of the art algorithms on the public MS challenge dataset.

Acknowledgment. The authors thank wholeheartedly the Grand Challenge 2008 organizers and J.C. Souplet and G. Malandain who kindly provided their pre-processed images of the MICCAI Challenge to allow a fair comparison of results. This work was partially funded by the Microsoft Prize, the CompuTumor INRIA funding and the German Academy of Sciences Leopoldina (Fellowship Programme LPDS 2009-10).

References

1. Leemput, K.V., Maes, F., Vandermeulen, D., Colchester, A.C.F., Suetens, P.: Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE Trans. Med. Imaging* 20(8), 677–688 (2001)

2. Dugas-Phocion, G., Ballester, M.Á.G., Malandain, G., Ayache, N., Lebrun, C., Chanalet, S., Bensa, C.: Hierarchical segmentation of multiple sclerosis lesions in multi-sequence MRI. In: ISBI, pp. 157–160. IEEE, Los Alamitos (2004)
3. Souplet, J.C., Lebrun, C., Ayache, N., Malandain, G.: An automatic segmentation of T2-FLAIR multiple sclerosis lesions. The MIDAS Journal - MS Lesion Segmentation (MICCAI 2008 Workshop) (2008)
4. Freifeld, O., Greenspan, H., Goldberger, J.: Multiple sclerosis lesion detection using constrained GMM and curve evolution. *J. of Biomed. Imaging* 2009, 1–13 (2009)
5. Akselrod-Ballin, A., Galun, M., Basri, R., Brandt, A., Gomori, M.J., Filippi, M., Valzasina, P.: An integrated segmentation and classification approach applied to multiple sclerosis analysis. In: CVPR 2006, pp. 1122–1129. IEEE, Los Alamitos (2006)
6. Menze, B.H., Kelm, B.M., Masuch, R., Himmelreich, U., Petrich, W., Hamprecht, F.A.: A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics* 10, 213 (2009)
7. Andres, B., Köthe, U., Helmstaedter, M., Denk, W., Hamprecht, F.A.: Segmentation of SBFSEM volume data of neural tissue by hierarchical classification. In: Rigoll, G. (ed.) DAGM 2008. LNCS, vol. 5096, pp. 142–152. Springer, Heidelberg (2008)
8. Shotton, J., Winn, J.M., Rother, C., Criminisi, A.: Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *Int. J. Comp. Vision* 81(1), 2–23 (2009)
9. Yi, Z., Criminisi, A., Shotton, J., Blake, A.: Discriminative, semantic segmentation of brain tissue in MR images. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 558–565. Springer, Heidelberg (2009)
10. Lempitsky, V.S., Verhoek, M., Noble, J.A., Blake, A.: Random forest classification for automatic delineation of myocardium in real-time 3D echocardiography. In: Ayache, N., Delingette, H., Sermesant, M. (eds.) FIMH 2009. LNCS, vol. 5528, pp. 447–456. Springer, Heidelberg (2009)
11. Criminisi, A., Shotton, J., Bucciarelli, S.: Decision forests with long-range spatial context for organ localization in CT volumes. In: MICCAI workshop on Probabilistic Models for Medical Image Analysis, MICCAI-PMMA (2009)
12. Styner, M., Lee, J., Chin, B., Chin, M., Commowick, O., Tran, H., Markovic-Plese, S., Jewells, V., Warfield, S.: 3D segmentation in the clinic: A grand challenge II: MS lesion segmentation. *MIDAS Journal*, 1–5 (September 2008)
13. Prima, S., Ayache, N., Barrick, T., Roberts, N.: Maximum likelihood estimation of the bias field in MR brain images: Investigating different modelings of the imaging process. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 811–819. Springer, Heidelberg (2001)
14. Rey, D.: Détection et quantification de processus évolutifs dans des images médicales tridimensionnelles: application à la sclérose en plaques. Thèse de sciences, Université de Nice Sophia-Antipolis (October 2002) (in French)
15. Prima, S., Ourselin, S., Ayache, N.: Computation of the mid-sagittal plane in 3d brain images. *IEEE Trans. Med. Imaging* 21(2), 122–138 (2002)
16. Evans, A.C., Collins, D.L., Mills, S.R., Brown, E.D., Kelly, R.L., Peters, T.M.: 3D statistical neuroanatomical models from 305 MRI volumes. In: IEEE-Nuclear Science Symposium and Medical Imaging Conference, pp. 1813–1817 (1993)
17. Amit, Y., Geman, D.: Shape quantization and recognition with randomized trees. *Neural Computation* 9(7), 1545–1588 (1997)
18. Breiman, L.: Random forests. *Machine Learning* 45(1), 5–32 (2001)
19. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Francisco (1993)