

Can we Distinguish Between Benign and Malignant Breast Tumors in DCE-MRI by Studying a Tumor’s Most Suspect Region Only?

Sylvia Glaßer, Uli Niemann, Bernhard Preim
Department for Simulation and Graphics
Univ. Magdeburg, Germany
glasser@isg.cs.uni-magdeburg.de

Myra Spiliopoulou
Knowledge Management and
Discovery Lab (KMD)
Univ. Magdeburg, Germany

Abstract

We investigate the task of breast tumor classification based on dynamic contrast-enhanced magnetic resonance image data (DCE-MRI). Our objective is to study how the formation of regions of similar voxels contributes to distinguishing between benign and malignant tumors. First, we perform clustering on each tumor with different algorithms and parameter settings, and then combine the clustering results to identify the most suspect region of the tumor and derive features from it. With these features we train classifiers on a set of tumors that are difficult to classify, even for human experts. We show that the features of the most suspect region alone cannot distinguish between benign and malignant tumors, yet the properties of this region are indicative of tumor malignancy for the dataset we studied.

1 Introduction

Breast cancer, i.e. malignant breast tumors, are often accompanied by neo-angiogenesis yielding an increased number of supporting blood vessels as well as an increased tissue permeability. In dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI), malignant tumor enhancement kinetics are often characterized by a rapid contrast agent washin and washout. For the evaluation of washin and washout, a region of interest (ROI) is defined and the ROI’s average contrast enhancement over time – the relative enhancement curve – is extracted. The radiologist assesses the early relative enhancement and the curve’s shape to analyze contrast agent washin and washout. Since breast tumors may consist of heterogeneous tissue parts, it is essential to detect the most malignant part for further diagnosis. Then, the tumor is rated as malignant as its most malignant part. Is it then feasible to automatically assess tumor malignancy by identifying and characterizing the tumor’s most suspect (i.e., malignant) region? In this study, we apply dif-

ferent clustering algorithms to breast tumors, identify the *Most Suspect Region* (MSRegion) for each clustering, and then use this region’s features to train classifiers, which exploit different combinations of features for learning. As a result, we identify which features contribute to the classification and should therefore be considered in clinical practice.

This study builds upon our previous work [8], where we studied how features of clusters derived from different clustering algorithms affect the discriminative power of a classifier. We extend that earlier work by learning classifiers on different sets of derived features, and by concentrating on the features of the most suspect region found through clustering, rather than the features of all clusters.

We use a dataset of 68 well chosen tumors (from 50 patients), of which 37 are malignant. This dataset comprises small tumors that are only detectable in DCE-MRI, cannot be classified with conventional mammography and are inherently difficult to separate, even for a human expert.

Our paper is organized as follows. In the next section, we describe related work on classification with features from the tumor regions. In Section 3, we first describe the DCE-MRI tumor data we analyzed and then elaborate on our learning method. The experimental results are presented in Section 4, and the findings are summarized in Section 5.

2 Related Work

The two major evaluation criteria for DCE-MRI breast tumors are the tumor’s morphology and its enhancement kinetics. We focus on the second criterion. Since benign tumors are predominantly more homogeneous than malignant ones [17], we expect a more heterogeneous enhancement with respect to their RE curves. Karahaliou et al. [12] analyzed the heterogeneity of breast tumors by evaluating the cross section of the largest tumor dimension. However, it is recommended to consider the whole tumor to improve the diagnostic accuracy [2]. To this purpose, a region merging method was adapted to breast DCE-MRI in [9], which

is used in [18] to predict malignancy based on the resulting amount and attributes of regions. Chen et al. [3] extract the most characteristic enhancement curve to differentiate between benign and malignant tumors.

ROI segmentation is a critical part of the breast cancer classification process. Marrocco et al. [15], who study microcalcifications on mammographic images, perform a watershed transform on the images. They input the ROIs' geometrical and textural features to decision-trees classifiers, which filter out uninteresting ROIs. The remaining regions are clustered into spatially contiguous clusters. We rather perform clustering before classification, and use the features of the most suspect region to classify the tumors. Nie et al. [16] focus on morphology/texture features of the MRI data to classify breast tumors, while Liang et al. [14] study the shape and margin features extracted from the minimum volume-enclosing ellipsoid of lesions that are segmented manually. Fusco et al. [7] also perform the segmentation manually, and then study different subsets of region features - dynamic ones, spatio-temporal ones and texture features. They showed that the subset of dynamic features achieves best performance. In our study, we also focus on dynamic features, but ROI segmentation is done automatically through clustering.

3 Materials and Method

In this section, we first describe the DCE-MRI breast tumor data set. Next, we give an overview of our method and elaborate on the tasks of clustering, region building and extraction of most suspect regions, feature specification and classification on selections of feature subsets.

3.1 Image Data

We study a set of 68 DCE-MRI breast tumors from 50 patients: 31 of the tumors are benign and 37 malignant; confirmation was carried out via histopathologic evaluation or by follow-up studies after six to nine months. We include only lesions that have been detected in MRI. The breast tumor data were acquired with a 1.0 T open MR scanner and are described in more detail in [8].

For each tumor voxel, the contrast enhancement is converted into relative enhancement (RE), i.e., the percent aged signal intensity increase, with

$$RE = (SI_c - SI) / SI \times 100 \quad (1)$$

Here, SI is the pre-contrast and SI_c is the post-contrast signal intensity. The RE plotted over time yields RE curves that allow for the extraction of the descriptive perfusion parameters (see Fig. 1(a)): washin (the steepness of the ascending curve), washout (the steepness of the descending

curve), peak enhancement (the maximum RE value), integral and time to peak (the time when peak enhancement occurs), which are substitutes for physiological parameters like tumor perfusion and vessel permeability. Since peak enhancement and integral strongly correlate, we exclude peak enhancement from the subsequent analysis.

The three-time-point (3TP) method presented by Degani et al. [5] allows for an automatic RE curve classification based on three well-chosen time points: t'_1 , the first point in time before the contrast agent injection, t'_2 , 2 min after t'_1 and t'_3 , 4 min after t'_2 . With the 3TP method, a RE change in the interval $\pm 10\%$ in the time between t'_2 and t'_3 will be interpreted as plateau, whereas RE changes higher than 10% and lower than -10% are classified as increasing curve and washout curve, respectively [5]. Since our study contains five - six time steps due to different scanning parameters, we assign the third time step to t'_2 and the last time step to t'_3 . The analysis of the initial contrast agent accumulation, i.e. the RE value at t'_2 , which is classified into slow, normal and fast in combination with the three curve shapes yields nine curve types (see Fig. 1(b)). We compare the results of our clustering algorithms with the 3TP classes.

3.2 Overview of the Classification Method

Our approach consists of three phases, see Fig. 2. In phase I, we perform several clustering runs with different parameter settings. For each run, we cluster together voxels that exhibit similar washin/washout curves (I.a). Then, we form spatially connected regions (I.b). Finally, we select the most suspect region *MSRegion*, or more specifically one *MSRegion* per good clustering (I.c). This step is described in Sec. 3.3.

In phase II, we derive a set of features that reflect the homogeneity of the region. Informally, a region is homogeneous if all its voxels are geometrically close to each other and exhibit the same washin/washout behavior. Hence, a

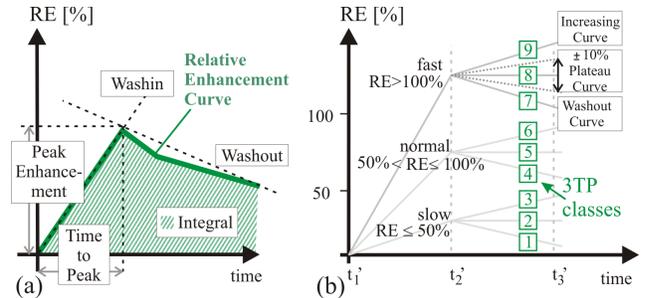


Figure 1. An RE curve and its perfusion parameters (a) and the 3TP classes based on RE at t'_1 , t'_2 , and t'_3 (b).

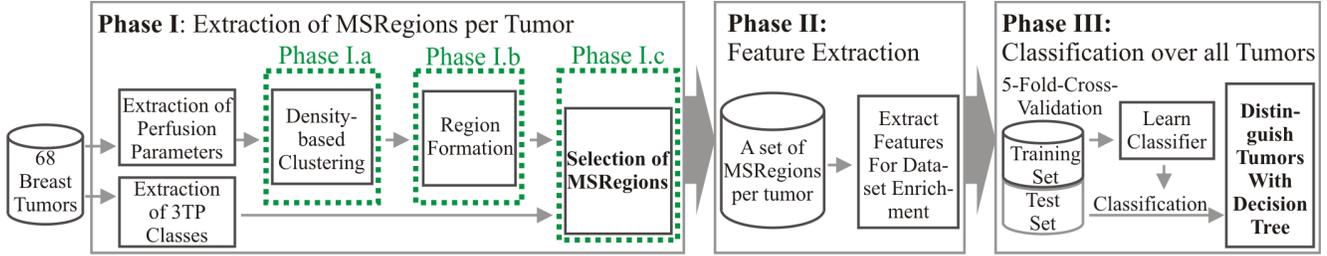


Figure 2. Schematic overview of the presented approach. In phase I, we perform several clustering runs (I.a) and form regions (I.b), of whom we select the MSRegions (I.c), see Sec. 3.3. Second, we extract features for each MSRegion for data enrichment (II). In phase III, we train learners over the set of all tumors and their MSRegions to distinguish between benign and malignant tumors.

region consisting only of 3TP class 8 voxels is more homogeneous than one with voxels from 3TP classes 7 and 8. Feature extraction is described in Sec. 3.4.

In phase III, we train classifiers on subsets of the features extracted from our data in phase II. Since phase I delivers one MSRegion per clustering run on the same tumor, we combine classifiers, each of which considers the MSRegion of a different run and predicts tumor malignancy (the label). A combination of such classifiers is an *ensemble*, which collectively decides on the label of a tumor by majority voting; each ensemble member has one vote. The specification of the classifiers is presented in Sec. 3.5.

3.3 Clustering and Identification of the Most Suspect Region

For the subsequently applied tumor classification, it is essential to form spatially connected and homogeneous voxel groups, i.e. with similar contrast enhancement kinetics. From these groups, we determine the MSRegion and derive the classification features later on.

First, we extract the four perfusion parameters washin, washout, time to peak, and integral per voxel (see Sec. 3.1) and carry out autoscaling. Second, we cluster on these parameters with a density-based clustering. We carry out Density-based Spatial Clustering of Applications with Noise (DBSCAN) [6], Density-Connected Subspace Clustering (SUBCLU) [11], and Ordering Points to Identify the Clustering Structure (OPTICS) [1]. The algorithms aggregate objects into clusters based on estimated density distributions. In contrast to other clustering algorithms, the density notation yields arbitrarily shaped clusters of any size, which is important for the underlying medical image data. Objects that do not exhibit similar parameter values to any cluster are marked as outliers. That’s a further advantage, since outliers may be caused by a missing inter-voxel-correspondence over time due to motion artefacts.

Each of the three algorithms requires two parameters:

the minimum number of points to build a cluster $MinPts$ and the maximum radius ϵ defining the neighborhood of a point. We set the values for the parameters as follows: $MinPts$ is empirically set to 4, 6 and 8. The calculation of ϵ depends on the clustering algorithm. For DBSCAN and each $MinPts$ value, ϵ was automatically determined as suggested in [4]. For SUBCLU, ϵ was extracted from the k-distances graph that maps the distance of an object to its k next neighbors (see also [8]). We apply this approach to all four perfusion parameter sets and assign ϵ to the mean of the four estimated values. For OPTICS, we experimentally determine ϵ by successively applying 20 potential ϵ values (in the range from 0.05 to 2) to the clustering algorithm. If a OPTICS run for one of the 20 values showed the best result compared to the 19 other ones, we increment the score of this ϵ value. To quantify the quality of a run we compare the sums of weighted values for Jaccard coefficient, F_1 score and Purity based on the actual clustering and the 3TP division. We pick the two best ϵ values. With $MinPts \in \{4, 6, 8\}$ we get three configurations for DBSCAN and SUBCLU and six configurations for OPTICS. To retain spatially connected clusters, each cluster’s voxels which are reachable within an 18-connected neighborhood are merged into groups.

Next, we determine the best clusterings by selecting the clusterings with the least outliers for DBSCAN, SUBCLU and OPTICS each. Thus, we yield the three best clustering runs: A_{DB} , A_{SC} , and A_{OP} . From the results of A_{DB} , A_{SC} , and A_{OP} , we reject all regions that contain less than three voxels. From the remaining regions, we choose the largest region with an average RE curve of 3TP class 7. If no such region exists, we search for the 3TP class 9, 8, 4, 6, 5, 1, 3, and 2 in that order. Although this is a user-defined ranking, we establish this empirical ranking based on definitions of the most malignant tumor enhancement kinetics: a present washout in combination with a strong washin. Thus, we obtain for each tumor and each clustering type (i.e., for A_{DB} , A_{SC} , and A_{OP}) three most suspect regions

$MSRegion_{DB}$, $MSRegion_{SC}$, and $MSRegion_{OP}$.

3.4 Feature Extraction of the MSRegion

The complete feature space describing a patient’s tumor is based on the MSRegion and the tumor attributes. We obtain the following sets of features: the features concerning the MSRegion’s size for each of the clusterings A_{DB} , A_{SC} , and A_{OP} (listed in Tab. 1), the features concerning homogeneity of these clusterings (listed in Tab. 2), and the features on washin and washout behavior based on the MSRegion and the corresponding clustering (listed in Tab 4). Features associated with the whole tumor are listed in Tab. 3

General features of MSRegion		
1	$\#RegVoxel$	Number of voxels of the MSRegion for a given clustering.
2	$RelRegSize$	The percentaged region size (with respect to the tumor size).

Table 1. Features of the MSRegion of a clustering; these features reflect the region size.

Features on Homogeneity and 3TP Class		
1	P	Purity value based on the comparison of the 3TP class division and a given clustering.
2	J	Jaccard coefficient based on the 3TP class division and a given clustering.
3	$F1$	F1-score value based on the 3TP class division and a given clustering.
4	$\#R_{3TP(7)}$	The number of regions with an average RE curve classified as 3TP class 7.

Table 2. Homogeneity features over the regions of a clustering.

Global Tumor Features		
1	$\#Voxel$	Number of tumor voxels.
2	age	Age of patient.

Table 3. Listing of global tumor features.

3.5 Specifying Classifiers and Ensembles

We define three base classifiers $L_1 - L_3$ for the three most suspect regions $MSRegion_{DB}$, $MSRegion_{SC}$, and $MSRegion_{OP}$. Next, we extract the intersection of these three regions resulting in the $MSRegion_{\cap}$, which is the starting point for the fourth base classifier L_4 . Furthermore, we

Washin and Washout Features for the MSRegion		
1	RE_{t_3}	The RE at the third time step.
2	$Integral$	The average value for the parameter integral.
3	$Washin$	The average value for washin.
4	$Washout$	The average value for washout.
5	TTP	The average value for time to peak.
6	$3TP_{region}$	The 3TP class of the region’s average RE curve.
7..15	$\#Voxels_{3TP(1..9)}$	Number of region voxels that have an 3TP curve 1..9.
16	$\#Present_{3TP}$	Number of different 3TP classes existing in the region.
17	$Majority_{3TP}$	3TP class with the most voxels in the region.

Table 4. Selection of washin and washout features for the MSRegion.

combine L_1-L_4 into two more classifiers: L_5 and L_6 . The detailed description of L_1-L_6 is listed in Tab. 5. Each of these classifiers learns a decision tree. We combine these classifiers into the ensembles $C_1(a) - C_1(d)$ described in the upper part of Tab. 6. Finally, we learn two baseline classifiers C_2 and C_3 that assign the label according to some intuitive but simplistic rule (cf. lower part of Tab. 6).

4 Experiments

In this section, we apply the learners and classifiers to our datasets to study the influence of the different clusterings and features. Our evaluation criterion is the number of *Correctly Classified Instances* $\#CCI$.

4.1 Experimental Settings

For the decision tree classifiers (Tab. 5), we employ the J4.8 classification algorithm (equivalent to the C4.5 algorithm [19]) of the *Waikato Environment for Knowledge Analysis* (Weka) library [10]. In the first run, we apply all complete attribute sets to the classification algorithm. Then, we try to find feature subsets of L_1-L_6 to optimize the number of $\#CCI$. Hence, we carry out two further runs (run 2(a) and (b)) and make use of the following methods (already available in Weka):

- a) The information gain of each attribute in all respective feature sets is evaluated. An attribute is excluded if its information gain is zero.

Decision Tree Classifiers	
L_1	We apply all features of $MSRegion_{DB}$ of Tab. 1-4.
L_2	We apply the features of $MSRegion_{SC}$ of Tab. 1-4.
L_3	We apply the features of $MSRegion_{OP}$ of Tab. 1-4.
L_4	We apply the features of $MSRegion_{\cap}$ (Tab. 4), all tumor attributes (Tab. 3) and region size attributes (Tab. 1).
L_5	We apply two features from each of the four MSRegions of $L_1 - L_4$: the region's size (Tab. 1(2)) and average 3TP class (Tab. 4(6)).
L_6	We extend L_5 . We use for each of the four MSRegions (of $L_1 - L_4$) the relative region size (Tab. 1(2)) and average 3TP curve (Tab. 4(6)). For the clustering with the least outliers from A_{DB} , A_{SC} , and A_{OP} , we extract the homogeneity measures (Tab. 2)(1)-(3). The tumor attributes are used (Tab. 3).

Table 5. Table of Learners $L_1 - L_6$.

- b) The wrapper approach, described in [13], is used to neglect irrelevant attributes. It delivers an attribute subset with the highest possible accuracy.

Since *age* and *#Voxel* show a high significance to almost every constructed decision tree, we then investigate the impact of these two global tumor attributes on the classification results. Therefore, we take the full attribute set and conduct three further runs where we exclude *age*, *#Voxel* and both, respectively. It becomes apparent that their influence on *#CCI* is very high. The restriction to attributes of the MSRegion only (run 3) results in a drastic decrease of maximum and average *#CCI*.

4.2 Results

For the first run, L_2 yields the best decision tree and correctly classifies 53 out of 68 ($\approx 78\%$) tumor instances. It uses the attributes *Majority_{3TP}*, *age* and *#Voxel*.

For the run 2(a), the maximum *#CCI* is 50 ($\approx 74\%$), achieved by L_2 , $C_1(b)$, and $C_1(d)$. L_2 is based on the $MSRegion_{SC}$. The decision tree of L_2 contains the attributes *J_{SC}*, *FI_{SC}*, *3TP_{region}*, *Majority_{3TP}*, and *age*. For run 2(b), L_2 also yields the best result with *#CCI* = 56 ($\approx 82\%$). This is the overall maximum of all runs as well. The corresponding decision tree (see Fig. 3(a)) consists of the attributes *Majority_{3TP}*, *age*, and *#Voxel*. Both runs 2(a) and 2(b), i.e., the information gain and the wrapper approach, yield an increased number of average *#CCI*.

The best tree that used solely MSRegion-based features

Ensemble Classifiers $C_1(a) - C_1(d)$	
$C_1(a)$ $C_1(b)$ $C_1(c)$ $C_1(d)$	The classifiers $C_1(a) - C_1(d)$ carry out a majority voting based on their inputs. If 50% of each classifier's inputs predict malignancy, then the classifier labels the tumor as malignant. We employ the following inputs (based on Tab. 5): $C_1(a)$ was created from $L_1 - L_6$, $C_1(b)$ was created from $L_1 - L_5$, $C_1(c)$ was created from $L_1 - L_4$ and L_6 , $C_1(d)$ was created from $L_1 - L_4$.
Rule-based Classifiers C_2 and C_3	
C_2	Labels the tumor based on the $MSRegion_{\cap}$. If its average curve type equals 3TP class 1, 4, or 7 the tumor is labelled as malignant. Otherwise it is labelled as benign.
C_3	If the number of voxels with an RE curve classified as 3TP class 1, 4, or 7 is greater or equal to 3, than the tumor is labelled as malignant. Otherwise it is labelled as benign.

Table 6. Table of Classifiers C_1 ($C_1(a) - C_1(d)$), C_2 , and C_3 based on $L_1 - L_6$ from Tab. 5.

was produced by L_4 . It classifies 45 ($\approx 66\%$) tumors correctly and is depicted in Fig. 3(b). The two baseline classifiers achieved a *#CCI* of no more than 38 ($\approx 56\%$). This means that simple rules of thumb are not adequate to distinguish between benign and malignant lesions on this difficult dataset.

5 Conclusion

With the presented approach, we examined the classification of DCE-MRI breast tumors based on their MSRegions. Therefore, we carried out density-based clustering to form groups of similar voxels with respect to their contrast enhancement. With the MSRegions from these clusterings, we derived features and focussed on the influence of the clustering, i.e., the feature set, to the classification. As a result, we showed that for the final classification step only a few features have to be employed. Another finding is the influence of patient age and tumor size. Both attributes strongly correlate with tumor malignancy and thus prove the challenges of automatic DCE-MRI breast tumor classification solely based on perfusion data. However, we identified features that correspond to tumor heterogeneity (i.e. Jaccard coefficient and Purity) and are also suited for distinguishing between benign and malignant tumors. In conclusion, the clustering and thus the grouping of the voxels into regions is more important for classification instead of the extraction

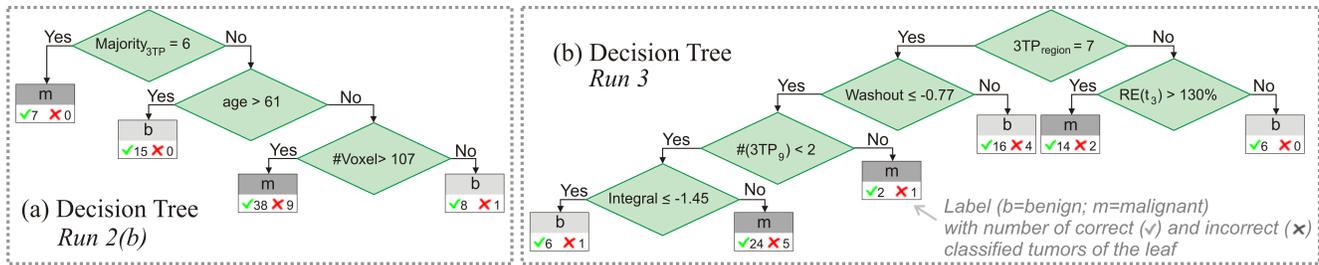


Figure 3. Decision tree (a) achieved the best #CCI value; decision tree (b) is the best among the classifiers that did not consider the features *age* and *#Voxel*.

of complex ensembles. We presented a feature set, which is best suited to automatically determine tumor malignancy.

For future work, we plan to apply our techniques to a bigger study. Furthermore, we want to investigate how the clinical expert could direct the clustering, i.e. set the clustering's parameters for a semi-automatic approach.

References

- [1] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander. OPTICS: Ordering Points To Identify the Clustering Structure. In *Proc. of ACM SIGMOD Conf. on Management of Data*, pages 49–60, 1999.
- [2] P. A. Baltzer, D. M. Renz, P. E. Kullnig, and et al. Application of computer-aided diagnosis (CAD) in MR-mammography(MRM). *Acad Radiol*, 16:435–442, 2009.
- [3] W. Chen, M. L. Giger, U. Bick, and G. M. Newstead. Automatic identification and classification of characteristic kinetic curves of breast lesions on DCE-MRI. *Medical Physics*, 33(8):2878–87, 2006.
- [4] M. Daszykowski, B. Walczak, and D. L. Massart. Looking for Natural Patterns in Data. Part 1: Density Based Approach. *Chemometrics and Intelligent Laboratory Systems*, 56:83–92, 2001.
- [5] H. Degani, V. Gusic, D. Weinstein, and et al. Mapping pathophysiological features of breast tumors by MRI at high spatial resolution. *Nat Med*, 3:780–782, 1997.
- [6] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proc. of Conf. on Knowledge Discovery and Data Mining (KDD)*, pages 226–231, 1996.
- [7] R. Fusco, M. Sansone, C. Sansone, and A. Petrillo. Segmentation and classification of breast lesions using dynamic and textural features in DCE-MRI. In *Proc. of Computer-Based Medical Systems (CBMS)*, pages 1–4, 2012.
- [8] S. Glaßer, U. Niemann, U. Preim, B. Preim, and M. Spiliopoulou. Classification of Benign and Malignant DCE-MRI Breast Tumors by Analyzing the Most Suspect Region. In *Proc. of Bildverarbeitung für die Medizin (BVM)*, page to appear, 2013.
- [9] S. Glaßer, U. Preim, K. Tönnies, and B. Preim. A visual analytics approach to diagnosis of breast DCE-MRI data. *Computers and Graphics*, 34(5):602 – 611, 2010.
- [10] G. Holmes, A. Donkin, and I. Witten. WEKA: a machine learning workbench. In *Proc. of the 2nd Australian and New Zealand Conf. on Intelligent Information Systems*, pages 357–361, 1994.
- [11] K. Kailing, H.-P. Kriegel, and P. Kröger. Density-Connected Subspace Clustering for High-Dimensional Data. In *Proc. of SIAM Conference on Data Mining*, pages 246–257, 2004.
- [12] A. Karahaliou, K. Vassiou, and N. S. e. a. Arikidis. Assessing heterogeneity of lesion enhancement kinetics in dynamic contrast-enhanced MRI for breast cancer diagnosis. *Br J Radiol*, 83(83):296–306, 2010.
- [13] R. Kohavi and G. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2):273–324, 1997.
- [14] X. Liang, K. Ramamohanarao, H. Frazer, and Q. Yang. A lesion shape and margin characterization method in dynamic contrast enhanced magnetic resonance imaging of breast. In *Proc. of Biomedical Imaging (ISBI)*, pages 1783 –1786, 2012.
- [15] C. Marrocco, M. Molinara, F. Tortorella, P. Rinaldi, L. Bonomo, A. Ferrarotti, C. Aragno, and S. Schiano lo Moriello. Detection of cluster of microcalcifications based on watershed segmentation algorithm. In *Proc. of Computer-Based Medical Systems (CBMS)*, pages 1–5, 2012.
- [16] K. Nie, J.-H. Chen, H. J. Yu, Y. Chu, O. Nalcioglu, and M.-Y. Su. Quantitative Analysis of Lesion Morphology and Texture Features for Diagnostic Prediction in Breast MRI. *Academic Radiology*, 15(12):1513–1525, 2008.
- [17] T. Okafuji, H. Yabuuchi, H. Soeda, Y. Matsuo, T. Kamitani, S. Sakai, M. Hatakenaka, S. Kuroki, E. Tokunaga, H. Yamamoto, and H. Honda. Circumscribed mass lesions on mammography: dynamic contrast-enhanced MR imaging to differentiate malignancy and benignancy. *Magn Reson Med Sci*, 7(4):195–204, 2008.
- [18] U. Preim, S. Glaßer, B. Preim, F. Fischbach, and J. Ricke. Computer-aided diagnosis in breast DCE-MRI – Quantification of the heterogeneity of breast lesions. *European Journal of Radiology*, 81(7):1532–1538, 2012.
- [19] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.