Computer-aided diagnosis in breast DCE-MRI—Quantification of the heterogeneity of breast lesions

Uta Preim\textsuperscript{a,∗}, Sylvia Glaßer\textsuperscript{b,1}, Bernhard Preim\textsuperscript{b,1}, Frank Fischbach\textsuperscript{c,2}, Jens Ricke\textsuperscript{c,2}

\textsuperscript{a} Herzzentrum Leipzig, Universitätsklinikum, Strümpellstraße 39, 04289 Leipzig, Germany
\textsuperscript{b} Otto-von-Guericke-Universität Magdeburg, FIN/ISG, Universitätsplatz 2, 39106 Magdeburg, Germany
\textsuperscript{c} Klinik für Radiologie und Nuklearmedizin der Universität Leipzig, Leipziger Straße 44, 39120 Magdeburg, Germany

\textbf{Abstract}

\textbf{Purpose:} In our study we aim at the quantification of the heterogeneity for differential diagnosis of breast lesions in MRI.

\textbf{Materials and methods:} We tested a software tool for quantification of heterogeneity. The software tool provides a three-dimensional analysis of the whole breast lesion. The lesions were divided in regions with similar perfusion characteristics. Voxels were merged to the same region, if the perfusion parameters (wash-in, wash-out, integral, peak enhancement and time to peak) correlated to 99%. We evaluated 68 lesions from 50 patients. 31 lesions proved to be benign (45.6%) and 37 malignant (54.4%). We included small lesions which could only be detected with MRI.

\textbf{Results:} The analysis of heterogeneity showed significant differences (\(p<0.005; \text{AUC 0.7}\)). Malignant lesions were more heterogeneous than benign ones. Significant differences were also found for morphologic parameters such as shape (\(p<0.001\)) and margin (\(p<0.007\)). The analysis of the enhancement dynamics did not prove successful in lesion discrimination.

\textbf{Conclusion:} Our study indicates that the region analysis for quantification of heterogeneity may be a helpful additional method to differentiate benign lesions from malignant ones.

© 2011 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Breast dynamic contrast enhanced MRI (DCE-MRI) offers the highest sensitivity for the detection of breast cancer. However, the specificity is problematic, because the enhancement kinetics of benign and malignant lesions often overlap\textsuperscript{[1]} and the morphologic features are often ambiguous\textsuperscript{[2]}. Therefore, the heterogeneity of breast lesions was evaluated to find an additional criterion to distinguish benign and malignant lesions. T1-weighted perfusion studies are used to observe the accumulation of contrast agent in the extravascular volume of the tissue\textsuperscript{[3,4]}. It has been shown that tumor growth depends on angiogenesis\textsuperscript{[5]} and the vascularity in tumors is often more permeable than in normal tissue\textsuperscript{[6]}. Malignant lesions show rim enhancement or heterogeneous enhancement due to necrosis and fibrosis mainly in the tumor center and angiogenetic activity predominantly at the periphery of the tumor\textsuperscript{[7]}. Benign tumors are predominantly more homogeneous\textsuperscript{[8]} and exhibit a lower vessel density than recurrent invasive breast cancer lesions\textsuperscript{[9]}. But fibroadenomas as well may display a heterogeneous internal enhancement due to mucinous or myxoid degeneration\textsuperscript{[10]}. However, such heterogeneous regions are most often found in larger tumors that are already known from mammography and ultrasound and histologically proven by ultrasound-guided biopsy.

Our hypothesis was that – due to angioneogenesis – malignant lesions are more heterogeneous than benign lesions, also when the lesion size is so small that necrosis not yet occurred. Other studies aimed at the quantification of heterogeneity. Karahaliou et al. analysed the heterogeneity and evaluated a cross section of the largest tumor dimension\textsuperscript{[11]}. However, it is recommended to consider the whole lesion to improve the diagnostic accuracy\textsuperscript{[12]}. To overcome the limitations of the two-dimensional analysis, we developed and tested a new software tool, the \textit{PERFUSIONANALYZER}, which enables the quantification of the heterogeneity of the whole lesion. The purpose of our study was to develop and test a software tool that quantifies the heterogeneity of small breast lesions. The aim was to evaluate and visualize small differences of the perfusion characteristics that are not perceptible in conventional MR images.
2. Materials and methods

The breast DCE-MRIs were clinically indicated. The patients gave their informed consent to the MRI exam and to the fact that the data might be used for research purposes. The retrospective study was performed according to the guidelines of the ethics committee.

2.1. Patients and lesions

We enrolled breast DCE-MRIs from January 2008 to December 2009. The database was collected in a retrospective fashion. We included datasets with lesions only detected in MRI. Palpable lesions or lesions detected in mammography or ultrasound were excluded. Datasets of 50 patients (mean age: 55, range: 36–73) showed 68 appropriate lesions. 31 lesions proved to be benign (45.6%) and 37 to be malignant (54.4%). The mean diameter was 8 mm (range: 4–18 mm). 60 lesions were confirmed by histopathology of specimens obtained by core needle biopsy. All biopsies were performed under MRI guidance with an MR-compatible fully automatic biopsy gun 100 mm, 14 G in vivo Germany. Eight lesions with MR-BI-RADS classification three proved to be benign by follow-up after six to nine months.

2.2. MRI protocol

The breast DCE-MRIs were performed on a 1.0 T open MR scanner (Philips Panorama HFO) with a dedicated breast coil (Philips Sense Breast). The imaging sequence was an axial T1 weighted 3D gradient echo sequence (TR 11, TE 6, flip angle 25°, FOV 300 × 320, voxel size 1.0 mm × 1.2 mm, matrix 320 × 267, slice thickness 3 mm without gap). Fat suppression was not employed. During and immediately after the bolus injection of contrast agent (0.1 mmol/kg body weight), one pre-contrast and four post-contrast images were acquired per series with a temporal resolution of 80 s. The contrast medium (Magnevist, Schering, Germany) was administered using an automated pump (Accutron MR MEDTRON 2007) with a flow of 1 ml/s. 16 of the 50 patients were premenopausal. In this subpopulation, the MR exam was performed in the 2nd week of the menstrual cycle.

2.3. Image interpretation and data analysis

The breast DCE-MRIs were evaluated by experienced radiologists. First, the datasets were examined at a commercially available workstation (Philips View Forum) using subtraction and parameter images in order to detect lesions unknown from ultrasound and X-ray-mammography. The radiologists assessed each lesion according to the MR-BI-RADS classification. The morphologic features (shape and margin) were evaluated and the lesion size was recorded. At this time, the histology of the lesions was unknown to the radiologists. We included lesions with MR-BI-RADS classification 3, 4 and 5.

2.4. Software tool

For further analysis of the depictsed lesions, the datasets were processed with our new software tool PERFFUSIONANLYZER (Fig. 1). The tool was developed with MeVisLab (MeVis Medical Solutions, Bremen). Motion correction was carried out with a combination of rigid and elastic registration [13]. The PERFFUSIONANLYZER displays several windows in parts known from conventional workstations: the source image, the subtraction image and several color-coded parameter images. This display enables the detection and segmentation of the lesion. The lesion was segmented semi-automatically. The radiologist used the subtraction image or the parameter image to depict the lesion and marked it with one click. After pressing the button “Get mask” on the right side of the PERFFUSIONANLYZER, the border of the lesion was calculated. The software program used the source images and explored the adjacent voxels of the marked seed point. Voxels with relative enhancement values ≥ 60% were considered to belong to the tumor and voxels < 60% to belong to the surrounding normal tissue, according to [1]. This limit was established for each lesion. However, the threshold applies only to the border voxels of the tumor. Voxels in the internal space could have values below this threshold, for instance due to necrosis. The radi-
The perfusion parameters were derived voxel-wise for the segmented lesions. We divided the lesion in regions with similar perfusion characteristics. Voxels were merged to regions if peak enhancement, wash-in, wash-out, time to peak and integral correlated to 99% (Pearson’s correlation). Regions must contain at least three voxels according to the recommendation for ROI setting [10]. Smaller regions were neglected, because regions of one or two voxels are predominantly influenced by noise and artefacts. Similar but not connected regions were considered as different regions. The region merging approach is explained in more detail in [14,15]. For quantitative analysis the averaged RECs of the regions were presented in a diagram. The curve presentation is only

Fig. 2. Perfusion parameters derived from REC.

Fig. 3. Benign lesion, (a) subtraction image, (b) segmented subtraction image, (c) region image, (d) REC view. The images show a lesion of 6 mm with oval shape and sharp margin, which was classified as MR-BI-RADS 3. 3TP classes 4, 5, 7 and 8 occurred. The lesion contained 3 regions suggesting homogeneity of a benign lesion. In follow-up MRIs the lesion did not change and was considered to be a fibroadenoma.
applied to regions with a minimum size of 10 voxels or 1% of the lesion size. The size of the tumor and the percentage size of each region (number of voxels) are provided in the diagram. The regions were color-coded according to the average RECs in the diagram. The region images and corresponding RECs are demonstrated in Figs. 3 and 4.

Furthermore, the curves were classified according to the 3 time point (3TP) method [16] in nine different classes (Fig. 5). For each lesion the number of the 3 time point classes was determined.

2.5. Statistics

We tested morphologic features, single perfusion parameters, numbers of regions and numbers of 3TP classes. The Mann–Whitney U test was applied to examine significant differences between benign and malignant lesions considering significance for \( p < 0.05 \). Receiver operating characteristic (ROC) curve analysis and calculation of the areas under the curve (AUC) were performed. We used SPSS for statistical analysis.

3. Results

Malignant lesions showed a significant higher number of regions compared to benign ones (median 17 versus 8, \( p = 0.005 \)). Therefore, malignant lesions were more heterogeneous. With a cut-off of ten regions sensitivity and specificity were 0.8 and 0.58, respectively. The lesion size of benign and malignant lesions was similar (mean 7.6 and 7.8 mm). The regions were color-coded and displayed in region images. The according REC of each region has the same color. Examples for a benign and a malignant lesion are given in Figs. 3 and 4.

The analysis of the 3TP classes revealed significant differences between the groups as well. Malignant lesions had more 3TP classes compared to benign lesions (mean 6.2 versus 5.3, \( p = 0.007 \)). With this method, the malignant lesions also turned out to be more heterogeneous. With a cut-off of five classes, sensitivity and specificity were 0.78 and 0.45, respectively.
Table 1
Median, mean, standard deviation, \( p \) (U-test).

<table>
<thead>
<tr>
<th></th>
<th>Benign</th>
<th></th>
<th></th>
<th>Malignant</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median</td>
<td>Mean</td>
<td>SD</td>
<td>Median</td>
<td>Mean</td>
<td>SD</td>
<td>Median</td>
<td>Mean</td>
</tr>
<tr>
<td>Patient age</td>
<td>61</td>
<td>57.5</td>
<td>10.5</td>
<td>54</td>
<td>53.7</td>
<td>8.3</td>
<td>0.032</td>
<td></td>
</tr>
<tr>
<td>Lesion size in mm</td>
<td>6</td>
<td>7.6</td>
<td>2.9</td>
<td>7</td>
<td>7.8</td>
<td>2.4</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>MR-BI-RADS</td>
<td>4</td>
<td>3.7</td>
<td>0.5</td>
<td>4</td>
<td>4.4</td>
<td>0.5</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Number of regions</td>
<td>8</td>
<td>13.6</td>
<td>29.5</td>
<td>17</td>
<td>25.3</td>
<td>24.3</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td>Number of 3TP classes</td>
<td>6</td>
<td>5.3</td>
<td>1.4</td>
<td>6</td>
<td>6.2</td>
<td>1.4</td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td>RE(^{a,b})</td>
<td>245.3</td>
<td>330.7</td>
<td>265.5</td>
<td>242.8</td>
<td>251.3</td>
<td>92.4</td>
<td>0.3</td>
<td></td>
</tr>
<tr>
<td>Wash-in(^{b})</td>
<td>120</td>
<td>152.3</td>
<td>125</td>
<td>117</td>
<td>121.7</td>
<td>48.3</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td>Wash-out(^{b})</td>
<td>−8.9</td>
<td>−8.5</td>
<td>25.3</td>
<td>−8.7</td>
<td>−9.4</td>
<td>12.7</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>Integral(^{b})</td>
<td>788.8</td>
<td>974.2</td>
<td>695</td>
<td>805.5</td>
<td>820.3</td>
<td>363.7</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>TTP(^{b})</td>
<td>3.2</td>
<td>3.7</td>
<td>1</td>
<td>3.2</td>
<td>3.5</td>
<td>0.56</td>
<td>0.8</td>
<td></td>
</tr>
</tbody>
</table>

SD, standard deviation; 3TP, 3 time point; TTP, time to peak.

\(^{a}\) Relative enhancement at the third time point.

\(^{b}\) The most suspect region.

Fig. 6. Receiver operating characteristic (ROC) curve of the criterion “number of regions”.

The perfusion parameters (peak enhancement, wash-in, wash-out, integral and time to peak) revealed no significant differences between the groups. The 3TP class 7 (strong early enhancement and following wash-out) occurred significantly more often in malignant lesions (18/37) than in benign ones (14/31). \( p \)-Value and AUC were 0.04 and 0.57, respectively. The sensitivity was determined to be 1.0, but the specificity was very poor (0.13). Figs. 6 and 7 show the ROC of the 3TP classes and of the region analysis. Fig. 8 demonstrates the box plots of the number of regions of benign and malignant lesions. Median, mean, standard deviation and \( p \)-values are displayed in Table 1. AUC values, sensitivity and specificity of selected cut-offs can be seen in Table 2.

Table 2
AUC, sensitivity and specificity for the significant parameters.

<table>
<thead>
<tr>
<th></th>
<th>AUC(^{a})</th>
<th>Cutoff</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of regions</td>
<td>0.7</td>
<td>10</td>
<td>0.8</td>
<td>0.58</td>
</tr>
<tr>
<td>Number of 3TP classes</td>
<td>0.68</td>
<td>5</td>
<td>0.78</td>
<td>0.45</td>
</tr>
<tr>
<td>3TP class 7</td>
<td>0.57</td>
<td>n.a.(^{a})</td>
<td>1.0</td>
<td>0.13</td>
</tr>
</tbody>
</table>

3TP, 3 time point.

\(^{a}\) Not applicable.

\(^{b}\) Area under curve.

Fig. 7. Receiver operating characteristic (ROC) curve of the criterion “number of 3TP classes”.

Despite the exclusive evaluation of lesions <2 cm, shape and margin were significantly different. Malignant lesions showed more often an irregular shape 22/37 (59.5%), but rarely a round 6/37 (16.2%), oval 8/37 (21.6%) or lobulated 1/37 (2.7%) shape. Benign lesions were more often characterized by a round shape 17/31 (54.8%), but rarely by an irregular 6/31 (19.4%), oval 4/31 (12.9%) or lobulated shape 4/31 (12.9%). The difference was highly significant (\( p < 0.001 \)). The margin showed significant differences as well (\( p = 0.007 \)). Malignant lesions showed more often a blurred margin 25/37 (67.6%) and rarely a well-defined margin 12/37 (32.4%). Benign lesions displayed more often a well-defined margin 21/31 (67.7%) and rarely a blurred margin 10/31 (32.3%). The histology of the lesion is shown in Table 3.

4. Discussion

Diagnostic criteria in breast DCE-MRI include assessment of morphological features like shape, margin and heterogeneity (internal architecture) as well as analysis of time intensity curves. Heterogeneity is known to be a feature of malignant tumors, because larger tumors develop necrotic areas [2]. Those necrotic areas can be estimated visually. However, in small lesions with-
Histology of the lesions.

4–9 occurred indicating suspect strong enhancement and follow-blurred margin and was classified as MR-BI-RADS 4. 3TP classes of 2 cm of the known tumor. The lesion had an irregular shape and images of a 45-year-old woman with new diagnosed breast cancer. change and was considered to be a fibroadenoma. Fig. 4 shows geneity of a benign lesion. In follow-up MRIs, the lesion did not and 8 occurred. The lesion contained 3 regions suggesting homoto[27]nicated.

In our study, the number of regions as well as the number of 3TP classes showed significant differences between benign and malignant lesions. Therefore, malignant lesions turned out to be more heterogeneous than benign ones. The region analysis revealed better results compared to the analysis of the 3TP classes. The reason could be that the 3TP classification has only nine classes, whereas the region analysis has much more regions and is thus more sophisticated.

Fig. 3 shows images of a 63-year-old woman who underwent mastectomy of the right breast. Breast DCE-MRI revealed another lesion of the left breast. The lesion of 6 mm with oval shape and sharp margin was classified as MR-BI-RADS 3. 3TP classes 4, 5, 7 and 8 occurred. The lesion contained 3 regions suggesting homogeneity of a benign lesion. In follow-up MRIs, the lesion did not change and was considered to be a fibroadenoma. Fig. 4 shows images of a 45-year-old woman with new diagnosed breast cancer. Breast DCE-MRI revealed an additional lesion of 6 mm in a distance of 2 cm of the known tumor. The lesion had an irregular shape and blurred margin and was classified as MR-BI-RADS 4. 3TP classes 4–9 occurred indicating suspect strong enhancement and follow-

Fig. 8. Box plots of the number of regions for benign and malignant lesions.

out apparent necrosis the visual analysis may not be sufficient. For that purpose, we developed a new software tool for quantification of heterogeneity, which enables the analysis of the whole lesion. Furthermore, the heterogeneity of small breast lesions has not yet been investigated.

In our study, the number of regions as well as the number of 3TP classes showed significant differences between benign and malignant lesions. Therefore, malignant lesions turned out to be more heterogeneous than benign ones. The region analysis revealed better results compared to the analysis of the 3TP classes. The reason could be that the 3TP classification has only nine classes, whereas the region analysis has much more regions and is thus more sophisticated.

Several commercially available software tools show colored parametric maps to display suspiciously enhancing lesions and map the RE and the curve type to colors. MTDynA (Mevis, Bremen, Germany) generates color parametric maps of relative changes in intensity of each voxel over time. Three time points are used to determine the color-coding: the pre-contrast, the first post-contrast and the final time point. Enhancement of the initial post-contrast time point relative to the baseline pre-contrast time point determines an initial color category. The final displayed colors (total of nine hues, three within each primary color category) for each voxel are determined by the rate of enhancement change between the initial post-contrast and last post-contrast points [17,18]. CADStream (Conﬁrma, Kirkland, WA) shows voxels with significant enhancement above a given threshold in color. The color encodes the post-initial change in the voxel values: wash-out is coded red, continuous enhancement blue and plateau green. It provides details about the lesions, like a synopsis of the full volume, the percentage of the tissue that shows wash-out, plateau and persistent enhancement. In addition, lesion identiﬁcation and enhancement distributions are presented. To obtain these features, the radiologist must draw a ROI, encompassing the whole lesion appropriately or must mark the lesion [19]. In the 3TP breast software package (CAD Sciences, White Plains, NY, USA), Hauth et al. employed the 3TP method for curve classiﬁcation. The ﬁrst time point is recorded before, the second 2 min and the third 6 min after contrast agent administration. The ﬁnal output includes color-hue- and color intensity-coded images. Initial enhancement rates are coded by color intensity. The post-initial enhancement changes are coded with three color hues: blue for increasing signal intensity, green for plateau, and red for wash-out [20]. However, these commercially available software tools are not able to quantify heterogeneity.

Several studies showed that the heterogeneity can be a helpful parameter [11,21–25]. Karahaliou et al. [11] created grey maps to visualize the heterogeneity, but they only applied the method to single perfusion parameters. They generated a map showing the heterogeneity of the initial enhancement, another map to visualize the post initial enhancement and one map of the signal enhancement ratio. In contrast, we combined five perfusion parameters to visualize the heterogeneity of the tumor. Furthermore, Karahaliou et al. analyzed only one slice of the tumor using the slice with the largest diameter. However, several studies suggest to analyze the whole lesion for taking the heterogeneity of the tumor into account [12,16]. In contrast to Karahaliou et al., our tool enables a three-dimensional analysis of the whole lesion. In our study, the analysis of the heterogeneity provided better results compared to the evaluation of the enhancement kinetics. The results suggest that the evaluation of heterogeneity is more robust and reliable.

Another advantage of the region approach is that manual ROI setting can be avoided. Though the above-mentioned software tools can guide manual ROI setting, they cannot avoid or replace it. Since the regions consist of voxels with similar perfusion parameters, partial volume averaging due to adjacent necrotic tissue or normal tissue is prevented.

Early strong contrast enhancement and following wash-out were not capable to predict dignity. Although the 3TP class 7 (early strong contrast enhancement and following wash-out) revealed significant differences between the groups with a sensitivity of 1.0, the speciﬁcity was very poor (0.13). This is in contrast to other studies [1,16,26]. However, a study of Williams et al. also failed to show significant differences between benign and malignant lesions [27]. The results of this study were explained by a long acquisition time.
and poor temporal resolution. In our study, the reason is probably the selection mode of the included lesions. We mainly included histologically proven lesions. That means the included lesions were suspicious enough to justify a biopsy. That way, typical benign lesions were not included. Another possible reason is, that with computer aided detection and a three-dimensional analysis voxels with suspect curve are detected without fail, which could be missed with the manual ROI method.

5. Study limitations

In our study, we excluded lesions without follow-up or histology. However, 8 lesions with the BI-RADS classification 3 had a follow-up only after 6–9 months. Due to the retrospective fashion and the limited period of our study, further follow-up with MRI or targeted ultrasound was not (yet) performed. However, this time frame is not sufficient to downgrade a lesion to BI-RADS 2. According to the BI-RADS lexicon a time frame of two years or longer is necessary.

Motion correction was carried out by the software tool and is essential to establish a valid inter-pixel correspondence over time, because breathing, heartbeat, patient movement and muscle relaxation can occur. The applied method represents the standard approach for motion correction of DCE-MRI data. Nevertheless, motion correction can also lead to artefacts like blurring and loss of small details. Furthermore, in some cases the motion artefacts cannot be completely removed.

The software is not optimized for speed which leads to long computation times for the analysis of larger lesions. In the future, a pre-processing step could solve this problem. In the pre-processing step, areas of no concern for the lesion could be removed. Another limitation is caused by the state of the software program, which is still a research prototype. Thus, the software consists of different not yet fully integrated parts, e.g. a part for segmentation and a separate part for heterogeneity analysis, and a careful workflow-adapted user interface is missing.

6. Conclusion

The three-dimensional quantification of the heterogeneity of small lesions detected in breast DCE-MRI showed significant differences between benign and malignant lesions. The analysis of heterogeneity was superior to the analysis of single perfusion parameters and enhancement curve types. Therefore, despite of an overlap of the number of regions in benign and malignant lesions, the quantification of the heterogeneity can be a helpful method in differential diagnosis and should be further evaluated.

Conflict of interest

The authors declare that they have no conflict of interest according to the guidelines of the International Committee of Medical Journal Editors.

Acknowledgement

This work was supported by the DFG in the framework of the Priority Programme 1335 “Scaleable Visual Analytics”.

References


