A simple and robust classification tree for differentiation between benign and malignant lesions in MR-mammography

Pascal A. T. Baltzer · Matthias Dietzel · Werner A. Kaiser

Abstract

Objectives In the face of multiple available diagnostic criteria in MR-mammography (MRM), a practical algorithm for lesion classification is needed. Such an algorithm should be as simple as possible and include only important independent lesion features to differentiate benign from malignant lesions. This investigation aimed to develop a simple classification tree for differential diagnosis in MRM.

Methods A total of 1,084 lesions in standardised MRM with subsequent histological verification (648 malignant, 436 benign) were investigated. Seventeen lesion criteria were assessed by 2 readers in consensus. Classification analysis was performed using the chi-squared automatic interaction detection (CHAID) method. Results include the probability for malignancy for every descriptor combination in the classification tree.

Results A classification tree incorporating 5 lesion descriptors with a depth of 3 ramifications (1, root sign; 2, delayed enhancement pattern; 3, border, internal enhancement and oedema) was calculated. Of all 1,084 lesions, 262 (40.4 %) and 106 (24.3 %) could be classified as malignant and benign with an accuracy above 95 %, respectively. Overall diagnostic accuracy was 88.4 %.

Conclusions The classification algorithm reduced the number of categorical descriptors from 17 to 5 (29.4 %), resulting in a high classification accuracy. More than one third of all lesions could be classified with accuracy above 95 %.

Key Points
• *A practical algorithm has been developed to classify lesions found in MR-mammography.*
• *A simple decision tree consisting of five criteria reaches high accuracy of 88.4%.*
• *Unique to this approach, each classification is associated with a diagnostic certainty.*
• *Diagnostic certainty of greater than 95% is achieved in 34% of all cases.*

Keywords Sensitivity and specificity · MR-mammography · Breast MRI · Classification tree · Decision tree

P. A. T. Baltzer (✉)
Department of Radiology, Medical University Vienna,
Währinger Gürtel 18-20,
1090 Vienna, Austria
e-mail: pascal.baltzer@meduniwien.ac.at

M. Dietzel
Department of Neuroradiology, University hospital Erlangen,
Erlangen, Germany

W. A. Kaiser
Institute of Diagnostic and Interventional Radiology 1,
University Hospital Jena, Jena, Germany

Introduction

Contrast-enhanced magnetic resonance imaging of the breast (MR-mammography, MRM) in comparison with mammography and ultrasound has high sensitivity for detection of breast cancer [1].

Malignant tumours show the phenomenon of neoangiogenesis, causing increased vasculature with pathological vessel architecture. Consequently, tumours strongly enhance after intravenous injection of contrast medium, reflecting pharmacokinetic properties of the tissue of interest. Absence of enhancement in malignant tumours is extremely rare resulting in a very high negative predictive value of MRM [2].

On the other hand, a variety of benign and malignant changes cause enhancement. Considering every enhancement as suspicious results in unnecessary biopsies; however, the number of such biopsies can be reduced by application of

simple diagnostic criteria [3]. Traditionally, dynamic enhancement patterns describing signal intensity changes over time have been used as diagnostic criteria [4, 5]. Malignant and benign hypervascularised lesions can thus be differentiated.

The known overlap between dynamic enhancement patterns between these lesion types makes the use of additional criteria mandatory. These have been mainly morphologic features, of which a multitude have been described. However, an increased number of diagnostic criteria does not always imply an increased diagnostic accuracy as different features may convey the same information. This has been shown for the features of the Breast Imaging Reporting And Data System (BI-RADS, [6]). Integration of morphologic and dynamic information into a diagnosis is a complex and experience-dependent task.

Therefore, objective classification schemes integrating morphological criteria are helpful in providing guidance for lesion differentiation. Several such schemes have been proposed [2, 7–12]. Limitations present in these studies were a lack of case numbers, lack of statistical validation, lack of feature selection, lack of colinearity consideration and, last but not least, lack of simplicity.

In clinical practice, an intuitive, transparent and simple classification algorithm is needed. Furthermore, a general estimation of diagnostic accuracy is impracticable; the reader and the referring clinician want to know what the specific combination of diagnostic criteria in a certain patient means. Integration of further information (i.e. high-risk or non-high risk patient, accidental finding or correlation with suspected pathology) can thus lead to individualised and flexible evidence-based diagnosis.

Classification trees are a suitable method for this purpose. A study population is homogenised by splitting using a hierarchical statistical selection of best diagnostic criteria [13]. This implies a selection of the most powerful diagnostic criteria, omitting criteria providing redundant information to those selected. The result is an intuitive classification tree where each possible combination of diagnostic criteria is associated with a specific predictive value or likelihood ratio. As a consequence, the reader knows to what degree his specific diagnosis is certain or not. The aim of this study was to provide such a simple and robust classification tree based on a representative large and coherent database.

Materials and methods

Patients and lesions

Consecutive patients examined with MRM at the University Hospital of Jena, Germany, over a time period of 12 years were eligible for this ethical review board-approved cross-sectional investigation. All patients provided written informed consent. MRI was performed because of unclear findings upon conventional imaging (BI-RADS 0, 3) or suspicious conventional findings (BI-RADS 4 and 5) or preoperative staging of biopsy-proven breast cancer (BI-RADS 6).

Inclusion criteria for the prospectively populated database used in this study were histopathology either by core biopsy or open surgery as reference standard after MRM. All histopathology workup was performed by board-certified breast pathologists of our university hospital's department of pathology in accordance with national S3 guidelines. Examinations after biopsy, surgery, chemotherapy or radiation therapy up to 1 year prior to MRI including BI-RADS 6 cases were excluded. Results from the same patient collective have been published in other contexts [14–23].

Imaging technique

All MRM examinations were performed in prone position using a field strength of 1.5 T and a dedicated double breast coil. The examination protocol included axial dynamic 2D gradient echo images with a temporal resolution of 1 min performed before and seven times after bolus injection of 0.2 mL Gd-DTPA (Magnevist, Bayer Healthcare, Leverkusen, Germany)/kg body weight at a flow rate of 3 mL/s, immediately followed by a 20-mL flush of saline solution. The injection procedure was carried out by a power injector. Subtraction of post-contrast from pre-contrast images was performed for fat saturation. Each examination was completed by a T2w TSE sequence. As a result of the timescale of this study, different MRI scanners were used. An MRI expert (W.A.K.) carefully supervised the examination protocol adjustment in order to yield consistent image contrast on different MR systems. Detailed protocols have been published previously (e.g. [14]).

Data analysis

Examinations were analysed by a consensus reading of two out of a pool of six radiologists blinded to histopathological results. Each radiologist had experience of more than 500 MRMs and all were trained by the same expert in MRM (W.A.K.).

Lesion size was assessed by means of electronic calipers and categorised into smaller than 0.5 cm, 0.5–1 cm, 1.1–2 cm, 2.1–3 cm, 3.1–5 cm, larger than 5 cm. Then, 16 further predefined categorical diagnostic criteria were assessed: initial enhancement (intermediate, strong), delayed enhancement curve type (persistent, plateau, washout), internal enhancement pattern (homogeneous, heterogeneous, centrifugal, centripetal/rim), blooming sign (present, absent), lesion shape (round, lobulated, irregular), signal intensity on pre-contrast non-fat-saturated T1w and T2w images (hyperintense,

isointense, hypointense, respectively), lesion margins (smooth, irregular), skin thickening (present, absent), nonenhanced internal structure (homogeneous, heterogeneous), destruction of nipple line (present, absent), vessels (absent, adjacent, prominent), internal septations (present, absent), hook sign (present, absent), root sign (present, absent), oedema (perifocal, diffuse ipsilateral, bilateral, absent) and lesion size. All have been published previously and described in detail [7, 14].

This categorical multivariate input data (17 variables) was used to construct a classification tree using Pearson chi-squared interaction detection (CHAID) methodology. The target variable was histopathology dichotomised into "benign" versus "malignant". Adjustments were as follows: minimum number of cases for parent nodes, $n=100$,;child nodes, $n=50$. The significance level for splitting nodes was set to $\alpha=0.05$ applying classical Bonferroni correction to avoid α error accumulation. Tenfold cross-validation was applied to assess the generalisability of the classification model. For estimation of general diagnostic accuracy, the area under the receiver operating characteristics (ROC) curve was calculated. Furthermore the accuracy of individual descriptor combination was assessed by standard measures. Here, the positive likelihood ratio (LR+) was chosen instead of predictive values, as it is independent from pretest probability [24].

Results

A total of 1,084 lesions in 1,012 patients (mean age 55.5, standard deviation 13.1 years) were analysed. Of these lesions, 648 (59.8 %) were malignant, showing 347 invasive ductal cancers, 108 invasive lobular cancers, 84 ductal carcinoma in situ (DCIS) and 109 invasive cancers not belonging to the aforementioned subgroups (i.e. mixed invasive ductal and lobular, invasive papillary, medullary and mucinous carcinoma). A total of 436 lesions were benign on histopathological analysis, revealing 103 fibroadenoma, 10 phylloid tumours, 83 papillomas, 220 proliferative fibrocystic changes and 20 inflammatory conditions. Median lesion size was 1.1–2 cm in each benign and malignant lesions and size ranged from smaller than 5 to larger than 5 cm in both subgroups (for size distribution, see Fig. 1).

The resulting CHAID tree demonstrated three ramifications with 10 terminal nodes, retaining 5 out of 17 initial criteria (cf. Table 1): margins (smooth vs. irregular), root sign (present vs. absent), oedema (diffuse ipsilateral or perifocal vs. absent or diffuse bilateral), internal enhancement pattern (inhomogeneous or centripetal (rim) vs. homogeneous or centrifugal) and delayed phase enhancement curve type (washout vs. plateau vs. persistent). All other criteria did not increase the accuracy of the classification algorithm. Tree details are shown in Fig. 2 and Table 2. Overall diagnostic accuracy (area under the ROC curve, AUC) was 0.884 (95 % CI 0.863–0.902, $P<0.0001$), cf Fig. 3.

The CHAID tree showed three terminal nodes with a classification accuracy greater than 95 % in 368 (34 %) of 1,084 lesions. A classification accuracy less than 75 % was identified in a minority of 196 (18.1 %) of 1,084 lesions, whereas the remaining 520 (48 %) lesions could be classified with an accuracy of 75.5–85.3 % (cf. Table 2).

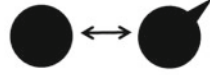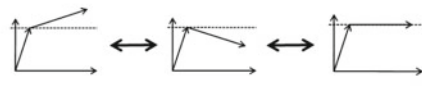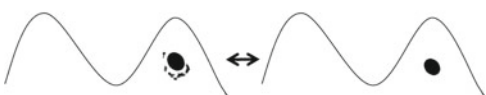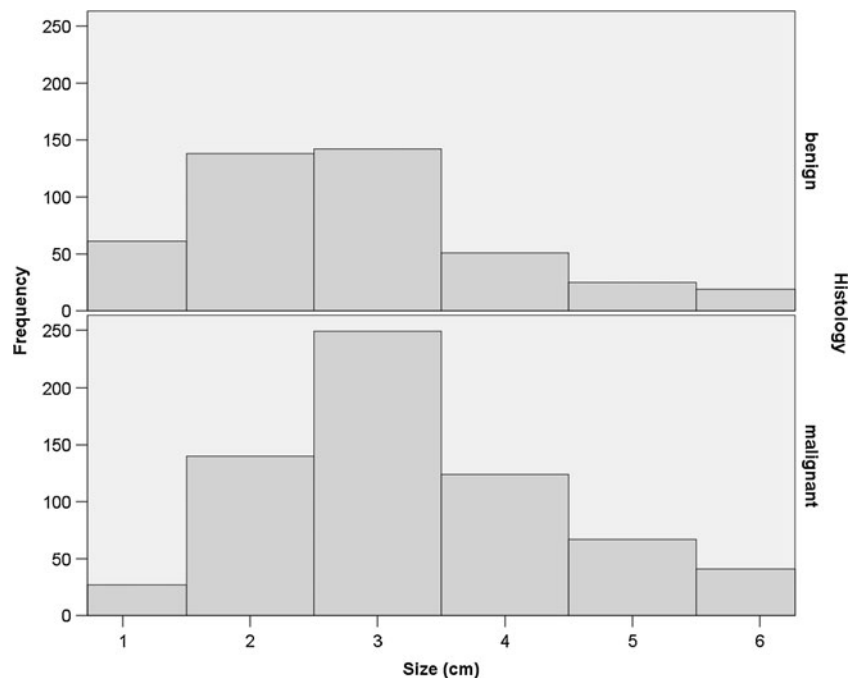Fig. 1 Size distribution in benign and malignant lesions: 1 <0.5 cm, 2 0.5–1 cm, 3 1.1–2 cm, 4 2.1–3 cm, 5 3.1–5 cm, 6 >5 cm

Table 1 Qualitative diagnostic criteria contained in the classification tree

Descriptor	Definition	Rating scheme
Root sign	Any spicule-like margin irregularity even in an otherwise smooth bordered lesion is considered as positive Root sign.	
Delayed enhancement	Relative signal difference between early and delayed phase are classified as follows: persistent signal increase as "Persistent", a steady signal as "Plateau" and a signal decrease as Washout.	
Margins	Margins can be either smooth or irregular on a general basis. Note that a sharply bordered lesion may present with a root sign.	
Internal enhancement pattern	Internal enhancement pattern is dichotomized by the CHAID algorithm. Inhomogeneous or centripetal (rim) enhancement hint at malignancy whereas homogeneous or centrifugal enhancement are indicative of a benign lesion.	
Edema	"Edema" refers to the presence of markedly prolonged T2 times of soft tissue not being caused by ductectasia. It causes a water like impression on T2w images and is a sign of malignancy if being ipsilateral of the lesion in perifocal or diffuse distribution. Absent or diffuse bilateral edema is rated as negative in the classification tree.	

Discussion

The result of our study is an easy, intuitive to follow and valid classification tree for differentiation between benign and malignant enhancing lesions in MRM. Using a simple table or tree scheme, one can classify lesions as being benign or malignant. Such diagnoses can be furthermore automatically classified into being "definite" or "most likely" or "likely to indeterminate" on the basis of the terminal nodes of the CHAID tree and corresponding LR+ value. Definite diagnosis with a diagnostic certainty of greater than 95 % could be achieved in 34 % of all cases. Only a minority of 18.1 % of cases showed a classification accuracy less than 75 %—and the reader knows when such a case is diagnosed. Such a statistically validated formal diagnostic decision algorithm providing the user with the diagnostic certainty of his diagnosis has not been described previously.

This may facilitate lesion classification in clinical practice: short-term follow-up and biopsy could be omitted in definitely benign lesions (i.e. high positive likelihood ratio), whereas a follow-up may be warranted in likely benign lesions (positive likelihood ratio clearly below 1). Although biopsy will always be mandatory for malignant lesions in order to decide whether to administer neoadjuvant chemotherapy, the classification tree may guide clinical decisions in cases of discrepant results between imaging and biopsy results. While a "likely malignant" diagnosis may be

explained by mastopathic changes, a "definite malignant" diagnosis in correlation with the same pathology would make further workup obligatory.

Indeterminate results defined as classification accuracy less than 75 % where found in only 18.1 % (196/1,084) of all lesions. This means that reasonable diagnosis is achieved for the majority of 81.9 % (888/1,084) lesions. Application of the classification tree does not take extra reading or computational time and may be specifically helpful to non-specialised readers.

Broad application is furthermore facilitated, as only simple dynamic and morphologic features were incorporated, avoiding the necessity for specific protocols or post-processing software.

Although quantitative lesion features would be preferable in order to eliminate subjective interpretation bias, measurement and reader bias still would be present in such techniques. The most commonly used quantitative MRI techniques in breast imaging are diffusion-weighted imaging (DWI), quantitative dynamic contrast-enhanced (DCE) MRI and proton spectroscopy [25]. Results of these techniques critically depend on technical acquisition parameters, limiting their use in clinical general classification schemes.

There is common ground that DCE and additional T2w imaging should be applied in every MRM [25]. As a consequence, our classification tree which was
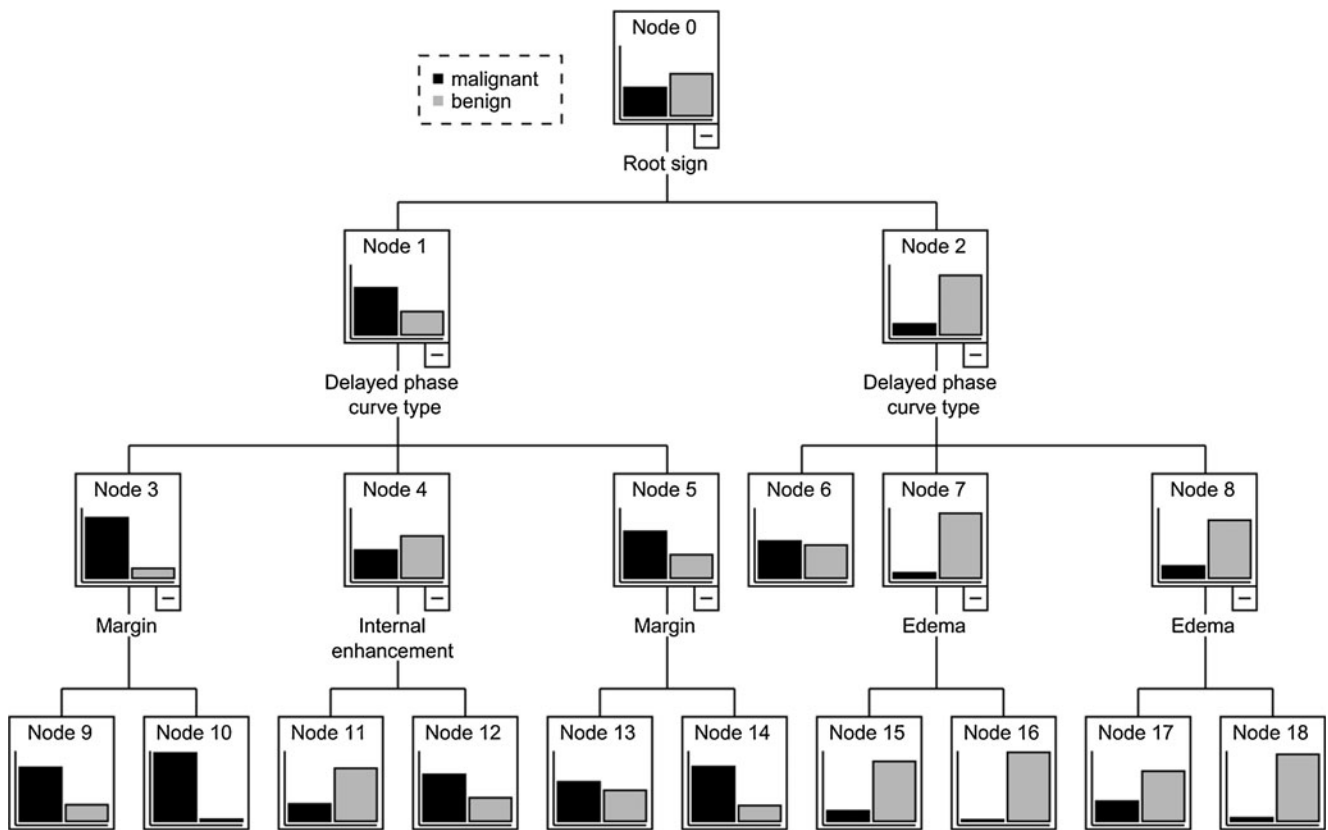
Fig. 2 Cross-validated chi-squared automatic interaction detection (CHAID) tree for differential diagnosis of benign (*grey*) vs. malignant (*black*) breast lesions. The initial study population (node 0, 648 malignant, 436 benign) is split into child nodes (node 1–18) by the independent variable showing highest discriminatory power based on chi-squared statistics. After 3 ramifications, the study population is split into 10 terminal nodes (node 9–18) where no further differentiation could be achieved (minimum node size was set to $n = 50$). In each node, *bars* indicate relative fraction of benign and malignant lesions with the *black line* on the *left* as the 100 % denominator. Nodes 16 and 18 have a positive predictive value for malignancy greater than 95 %, whereas node 10 has a negative predictive value for malignancy of 97.3 %. Detailed node results are given in Table 2

developed on the basis of such a protocol is generally applicable in clinical practice. Of note, the positive predictive values given in the terminal nodes of our tree are only valid for the prevalence of malignancy (59.8 %) in our patient group as predictive values mathematically depend on pretest probability, i.e. prevalence. That is why we generally prefer likelihood ratios for all tree nodes as a more appropriate statistical means. The reader can thus convert the individual pretest probability of malignancy into a post-test probability. This approach is highly flexible and allows integration of further patient-related data like prior imaging results, family and personal history. Our example cases illustrate this (Figs. 4, 5 and 6). Of particular interest in this context is case 3 (Fig. 6), which was not part of our study, as no histopathological tissue sampling was obtained. This was due to the fact that it was a 48-year-old woman presenting with an architectural distortion on mammography and ultrasound of the left breast which presented as nonenhancing fibrotic area on MRI, consistent with the diagnosis of mastopathy; the non-mass enhancement was an incidental finding without any clinical,

mammographical or ultrasound correlate. Our general prevalence of malignancy per individual in MRM examinations due to regular quality controls is less than 30 %. Assuming a maximum cancer prevalence of 30 %, the positive predictive value for malignancy is 8.1 %. After informational conversation with the patient, we chose to refrain from taking a biopsy. Long-term follow-up of 5 years confirmed a benign finding, most likely due to mastopathic changes. Lesion workup would have been completely different if the lesion showed a correlate with suspicious clustered microcalcifications, strongly increasing the probability of malignancy in this case. Our example case underlines that a classification algorithm should always be applied in a flexible manner in order to assist but not replace the radiologist's diagnosis.

Our study has potential limitations. Firstly, only histopathologically verified lesions were considered, potentially biasing our lesion database towards more difficult diagnosis in which invasive diagnosis was requested in the first place. We did so in order to provide a more accurate reference standard. Furthermore, the selection bias towards higher prevalence of

Table 2 Detailed node characteristics of the classification tree

Node	Definition	Malignant, n (%)	Benign, n (%)	LR+ (95 % CI)
1	Root sign (−)	172 (33 %)	349 (67 %)	0.33 (0.29–0.38)
2	Root sign (+)	476 (84.5 %)	87 (15.5)	3.68 (3.03–4.47)
3	Root sign (−), persistent (+)	32 (13.9 %)	198 (86.1 %)	0.11 (0.08–0.15)
4	Root sign (−), washout (+)	97 (59.9 %)	65 (40.1 %)	1 (0.75–1.34)
5	Root sign (−), plateau (+)	43 (33.3 %)	86 (66.7 %)	0.34 (0.24–0.48)
6	Root sign (+), persistent (+)	32 (47.1 %)	36 (52.9 %)	0.6 (0.38–0.95)
7	Root sign (+), washout (+)	336 (92.3 %)	28 (7.7 %)	8.07 (5.6–11.64)
8	Root sign (+), plateau (+)	108 (82.4 %)	23 (17.6 %)	3.16 (2.05–4.87)
9	Root sign (−), persistent (+), margin (irregular)	29 (23.4 %)	95 (76.7 %)	0.21 (0.14–0.31)
10	Root sign (−), persistent (+), margin (smooth)	3 (2.8 %)	103 (97.2 %)	0.02 (0.01–0.06)
11	Root sign (−), washout (+), intensity enhancement (inhomogeneous, rim)	77 (75.5 %)	25 (24.5 %)	2.07 (1.34–3.2)
12	Root sign (−), washout (+), intensity enhancement (homogeneous, centrifugal)	20 (33.3 %)	40 (66.7 %)	0.34 (0.2–0.57)
13	Root sign (−), plateau (+), margin (irregular)	29 (43.9 %)	37 (56.1 %)	0.53 (0.33–0.84)
14	Root sign (−), plateau (+), margin (smooth)	14 (22.2 %)	49 (77.8 %)	0.19 (0.11–0.34)
15	Root sign (+), washout (+), oedema (absent, bilateral)	139 (85.3 %)	24 (14.7 %)	3.9 (2.57–5.91)
16	Root sign (+), washout (+), oedema (perifocal, ipsilateral)	197 (98 %)	4 (2 %)	33.14 (12.41–88.51)
17	Root sign (+), plateau (+), oedema (absent, bilateral)	50 (71.4 %)	20 (28.6 %)	1.68 (1.02–2.78)
18	Root sign (+), plateau (+), oedema (perifocal, ipsilateral)	58 (95.1 %)	3 (4.9 %)	13.01 (4.1–41.25)

Tree structure and corresponding node numbers are in accordance with Fig. 2. LR+ denotes positive likelihood ratio if node is rated positive for malignancy. Nodes 1 and 2 represent 1st ramification, 3 to 8 the second ramification and 9 to 18 the 3rd ramification (terminal nodes)
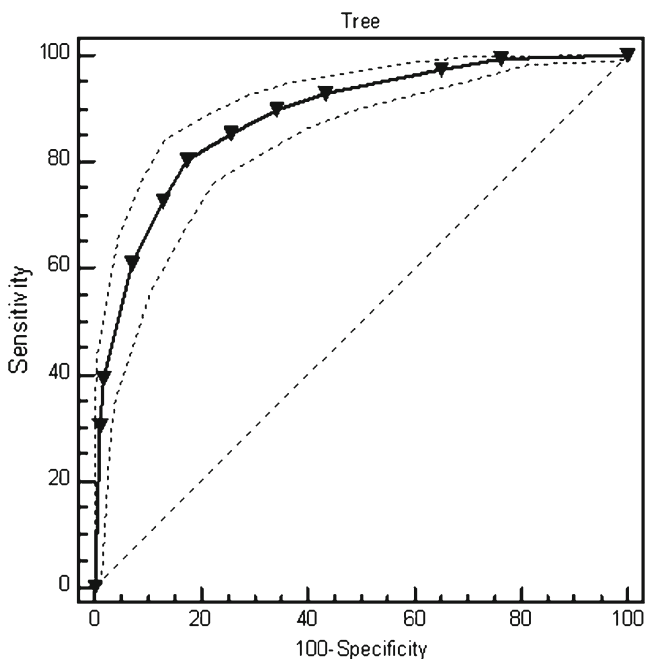

Fig. 3 Receiver operating characteristics (ROC) curve (*black line*) of tenfold cross-validated CHAID tree results. *Dotted lines* represent 95 % confidence interval margins. The area under the curve is calculated as 0.885, standard error 0.00995, 95 % confidence interval 0.864 to 0.903. Probability of malignancy increases in the node order 10, 14, 9, 12, 13, 17, 11, 15, 18, 16 (cf. Fig. 2 and Table 2)

malignancy is addressed by giving prevalence-independent statistical results. This is a significant difference to previous work in this field. Second, no quantitative techniques were used. Recommendations for MRM include dynamic and T2-weighted protocols, but not DWI, quantitative DCE or spectroscopy because the diagnostic value of these techniques has not been clarified yet. Furthermore, results of quantitative MRI techniques critically depend on the acquisition parameters. More complex classification algorithms could include more data—however, at the cost of simplicity and robustness. It should also be kept in mind that a dynamic plus T2w only protocol is much shorter than a multiparametric MRI protocol. Finally, our lesion features are not totally congruent with the BI-RADS lexicon because our prospectively populated database was started prior to the publication of the BI-RADS lexicon. Furthermore, the BI-RADS lexicon is not established in all countries, e.g. the UK [26]. Finally, the BI-RADS criteria are of limited use in non-mass enhancements (such as our case 3) and have not been shown to be more valid or reliable than other diagnostic features. No formal classification algorithm is provided in the BI-RADS lexicon. By contrast, the diagnostic content of the BI-RADS criteria was not validated in a multivariate analysis. Furthermore, colinearity of BI-RADS criteria has been described, meaning that there is an overlap in the inherent diagnostic information of these criteria [6]. The results published by Demartini and others hint at the necessity of reconsidering BI-RADS criteria

Fig. 4 Example case of a 76-year-old woman presenting with a mass lesion in the right breast at 9 o'clock. **a** Pre-contrast T1w, **b** early enhanced phase, **c** delayed enhanced phase, **d** T2w-TSE. Classification according to the CHAID tree is as follows: root sign positive (node 2), washout in delayed phase enhancement (node 7), positive perifocal oedema (terminal node 16, cf. Fig. 2). Resulting probability of malignancy based on the database of this paper is 98 %, prevalence-independent LR+ is 33.2, meaning that the initial probability of malignancy is increased 33-fold. Core biopsy showed invasive ductal carcinoma G2
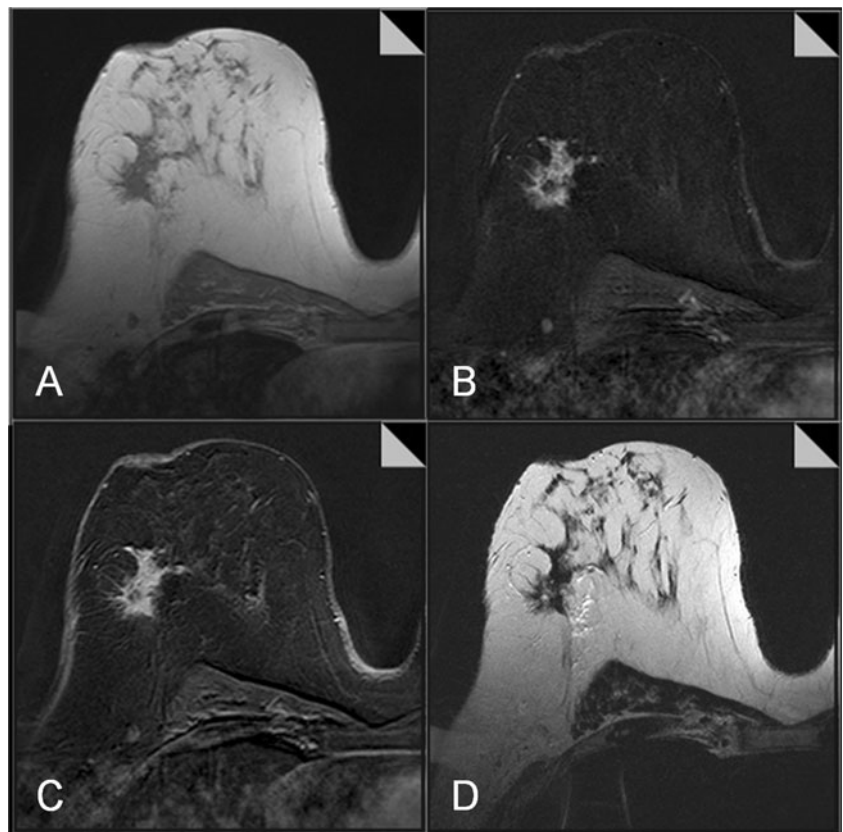

Fig. 5 Example case of a 28-year-old woman showing a mass lesion in the left breast at 12 o'clock. **a** Pre-contrast T1w, **b** early enhanced phase, **c** delayed enhanced phase, **d** T2w-TSE. Classification according to the CHAID tree is as follows: negative root sign (node 1), persistent delayed phase enhancement (node 3), smooth margins (terminal node 10, cf. Fig. 2). Resulting probability of malignancy based on the database of this paper is 2.8 %, prevalence independent LR+ is 0.02, meaning that the initial probability of malignancy is decreased 50-fold. Core biopsy showed benign fibroadenoma
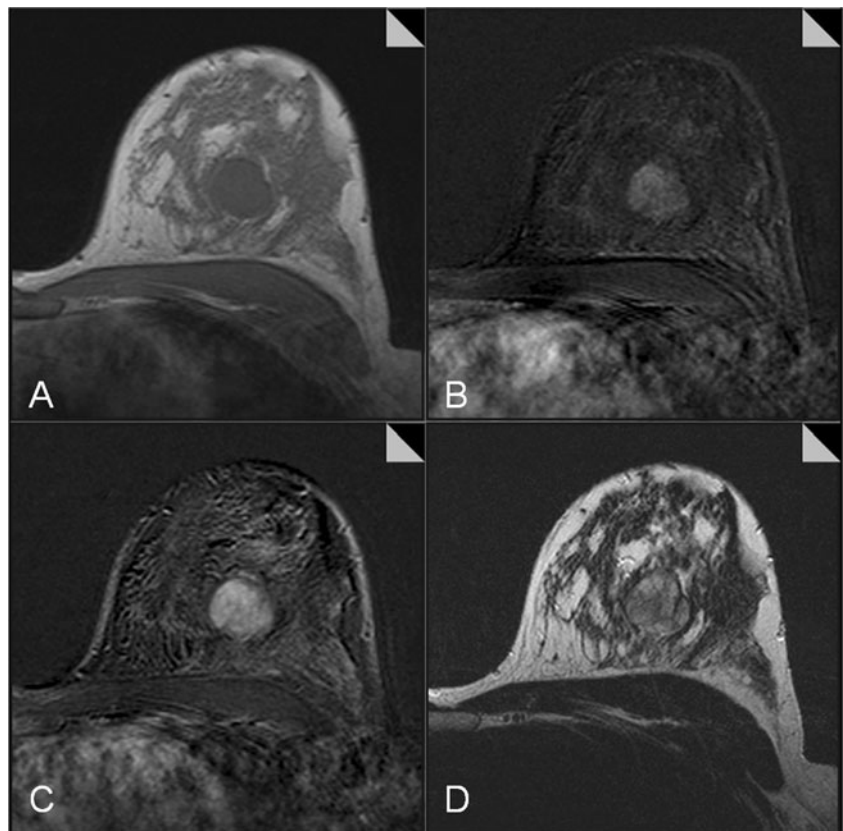
Fig. 6 Example case of a 48-year-old woman showing an incidental non-mass enhancement in the lower right breast at 6 o'clock (not from the database of this study). **a** Pre-contrast T1w, **b** early enhanced phase, **c** delayed enhanced phase, **d** T2w-TSE. Classification according to the CHAID tree is as follows: negative root sign (node 1), persistent delayed phase enhancement (node 3), irregular margins (terminal node 9, cf. Fig. 2). Resulting probability of malignancy based on the database of this paper is 23.4 %, prevalence independent LR+ is 0.2, meaning that the initial probability of malignancy is reduced by five times. The patient had a negative family history of breast cancer. Consequently, 5-year follow-up including MRI confirmed a benign finding
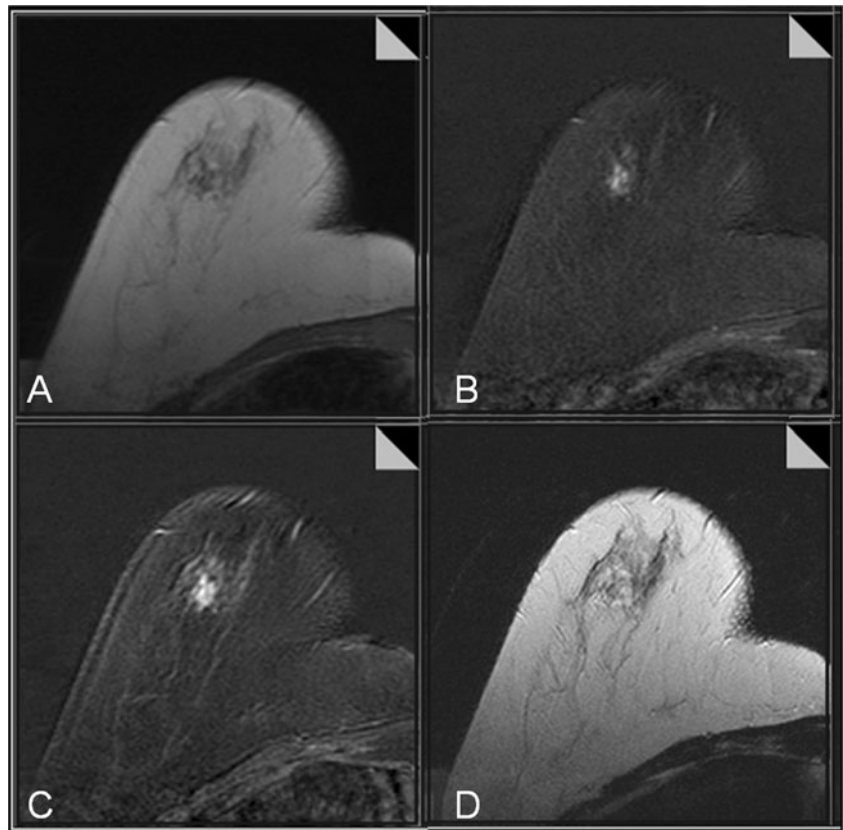
according to their statistically validated diagnostic use [11, 27, 28]. We did not systematically address the reproducibility of our results. While the cross-validation performed in this study is a sufficient method to prove the robustness of the proposed classification model considering the high number of cases analysed, a prospective study could give insights into the applicability of the model in clinical practice. The consensus reading approach used in our study does not allow assessment of reproducibility of the diagnostic criteria incorporated into the model. According to our experience, no significant disagreement in assessment of diagnostic criteria was observed during data analysis in this study. This may be due to the simplicity of the criteria used.

Our study is not the first to propose a classification algorithm in MRM. Nunes et al. initially presented a classification tree approach in a split sample of 192 patients [8]. The authors included 10 features including the criterion enhancement vs. no enhancement. No dynamic enhancement features were used. The statistical parameters for model construction were not given in the article and some effect of overfitting due to low sample size is likely. Lesions not palpable or not visible by mammography were omitted, resulting in a validated accuracy of 83 %. Our own approach incorporates five criteria including the dynamic curve type and reaches a cross-validated accuracy of 88.4 % based on inclusion of all enhancing histopathologically verified lesions independent of clinical or mammographic findings.

A more simple approach has been proposed by the Göttingen group in 2002 [2]. This score is completely empirical and includes five diagnostic criteria, two of which describe initial and delayed dynamic enhancement patterns. A scoring sheet is used to classify lesions on the basis of a BI-RADS analogous five-point confidence scale. Prevalence of malignancy was 50.6 % in 265 investigated enhancing lesions; only 8 of 136 (5.9 %) of all cancers were DCIS. Sensitivity and specificity of 92 % were reached. It is not stated whether the classification scheme was constructed on basis of the same data it was tested on. Consequently the Göttingen method achieved lower diagnostic accuracy (sensitivity 83.1 % and specificity of 58.8 %) when used by another group [29]. By contrast, our classification model provides diagnostic certainty for all subcategories, is more intuitive to follow and is based on a larger and more representative database (i.e. more DCIS cases). A model describing the criteria investigated in this study included a variety of mainly morphologic features [7]. Simple scores were calculated on the basis of feature prevalences in benign and malignant lesions. Colinearity was not addressed and feature selection was not performed. The model was validated in a group of 132 histologically verified lesions and revealed a sensitivity and specificity of 90.9 % and 60 %, respectively. The authors suggested to combine the Göttingen model with the extended scoring system, achieving a sensitivity of 97 % and a specificity of 76.5 % [29].

Considering mass lesions only, Tozaki et al. described an empirical statistically non-validated classification algorithm having a sensitivity of 100 % and a positive predictive value of 98 %. The pretest probability of malignancy in their group was 49 of 63 lesions (77.7 % [10]). The same group published a similar interpretation model for non-mass lesions based on 30 non-mass lesions (18 malignant, prevalence of malignancy 60 %). A positive predictive value of 94 % was reported [9]. Application of this model to a group of 102 lesions (10 malignant) could not reproduce this high predictive value [30]. Low sample size and patient selection limit the applicability of this classification system. A comprehensive approach based on data of 995 conventionally BI-RADS 4 and 5 lesions obtained in a multicentre study by 27 interpreting radiologists was published by Schnall and co-workers in 2006 [12]. A number of protocols were used, some with high temporal resolution dynamic imaging. The interpretation model was restricted to mass lesions only and consisted of a combination of CART algorithm and logistic regression and finally selected 10 features also including non-MRI features of age, whether the lesion was palpable and whether microcalcifications were present. There was no split sample or cross-validation performed. In a direct comparison, our model is simpler and intuitive, cross-validated and was achieved by consistent protocols and is based on image interpretation by radiologists trained in the same institution in order to optimise reliability. Another interpretation model including qualitative enhancement characteristics, internal enhancement pattern and lesion margin characteristics was published in 2006 [26]. The model was constructed on the basis of patient data mainly from a high-risk screening study mixed with 100 symptomatic women (prevalence of malignancy 86 out of 991, 8.7 %). Logistic regression analysis was the chosen statistical method and the model was validated in a split sample approach reaching an AUC of 0.88. It should be kept in mind that the low cancer prevalence or high number of true negative findings suggests a higher diagnostic accuracy as compared to studies with histopathological verification and, consequently, a higher prevalence of malignancy. A more recent single-centre study investigated multivariate models for prediction of malignancy in 855 lesions (155, 18.1 % malignant) detected in cancer staging or high-risk screening [11]. The best predictive model identified considered three criteria (indication, size and kinetics) and reached an AUC of 0.7. Morphologic criteria as included in prior predictive models [2, 7–10, 12] were less important [11]. By contrast, our own model is independent of lesion size and mainly relies on morphologic criteria. A likely explanation for our observation is that the specific criteria retained by our classification tree are not standard descriptors (e.g. root sign, oedema), whereas "classical" BI-RADS descriptors have been reported to be of limited diagnostic potential in previous studies [6, 11, 27, 28].

In summary, a variety of different interpretation models have been proposed. All contain delayed enhancement curve type (persistent, plateau, washout). Border characteristics and internal enhancement pattern seem to be the most important morphologic criteria [2, 7]. It is assumed that variability of patient characteristics, diagnostic criteria used and limited statistical validation is a major reason for the variety of results of interpretation models published so far. Predictive values depend on pretest probability and should thus be considered with care in different clinical contexts. This is why we provide likelihood ratios which allow one to model individual results [24].

In conclusion, we provide a simple and robust classification model which follows an intuitive tree structure reflecting a structured step by step diagnostic process for lesion differentiation. It is based on the largest database published so far in this context, thus ensuring statistical stability. Every single combination of diagnostic criteria is associated with a specific and relative likelihood of malignancy and thus provides the reader with an objective tool for making clinical decisions.

References

1. Warner E, Messersmith H, Causer P, Eisen A, Shumak R, Plewes D (2008) Systematic review: using magnetic resonance imaging to screen women at high risk for breast cancer. Ann Intern Med 148:671–679
2. Baum F, Fischer U, Vosshenrich R, Grabbe E (2002) Classification of hypervascularized lesions in CE MR imaging of the breast. Eur Radiol 12:1087–1092
3. Williams TC, DeMartini WB, Partridge SC, Peacock S, Lehman CD (2007) Breast MR imaging: computer-aided evaluation program for discriminating benign from malignant lesions. Radiology 244:94–103
4. Kaiser WA, Zeitler E (1989) MR imaging of the breast: fast imaging sequences with and without Gd-DTPA. Preliminary observations. Radiology 170:681–686
5. Kuhl CK, Mielcareck P, Klaschik S et al (1999) Dynamic breast MR imaging: are signal intensity time course data useful for differential diagnosis of enhancing lesions? Radiology 211:101–110
6. Benndorf M, Baltzer PAT, Kaiser WA (2011) Assessing the degree of collinearity among the lesion features of the MRI BI-RADS lexicon. Eur J Radiol 80:e322–e324
7. Malich A, Fischer DR, Wurdinger S et al (2005) Potential MRI interpretation model: differentiation of benign from malignant breast masses. AJR Am J Roentgenol 185:964–970
8. Nunes LW, Schnall MD, Orel SG et al (1997) Breast MR imaging: interpretation model. Radiology 202:833–841

9. Tozaki M, Fukuda K (2006) High-spatial-resolution MRI of non-masslike breast lesions: interpretation model based on BI-RADS MRI descriptors. AJR Am J Roentgenol 187:330–337

10. Tozaki M, Igarashi T, Matsushima S, Fukuda K (2005) High-spatial-resolution MR imaging of focal breast masses: interpretation model based on kinetic and morphological parameters. Radiat Med 23:43–50

11. Demartini WB, Kurland BF, Gutierrez RL, Blackmore CC, Peacock S, Lehman CD (2011) Probability of malignancy for lesions detected on breast MRI: a predictive model incorporating BI-RADS imaging features and patient characteristics. Eur Radiol 21:1609–1617

12. Schnall MD, Blume J, Bluemke DA et al (2006) Diagnostic architectural and dynamic features at breast MR imaging: multicenter study. Radiology 238:42–53

13. Kass GV (1980) An exploratory technique for investigating large quantities of categorical data. Appl Stat 29:119

14. Dietzel M, Baltzer PA, Vag T et al (2010) Differential diagnosis of breast lesions 5 mm or less: is there a role for magnetic resonance imaging? J Comput Assist Tomogr 34:456–464

15. Dietzel M, Baltzer PAT, Vag T et al (2010) Application of breast MRI for prediction of lymph node metastases - systematic approach using 17 individual descriptors and a dedicated decision tree. Acta Radiol 51(8):885–894

16. Dietzel M, Baltzer PA, Vag T et al (2010) Magnetic resonance mammography of invasive lobular versus ductal carcinoma: systematic comparison of 811 patients reveals high diagnostic accuracy irrespective of typing. J Comput Assist Tomogr 34(4):587–595

17. Dietzel M, Baltzer PA, Vag T et al (2011) Potential of MR mammography to predict tumor grading of invasive breast cancer. Rofo 183(9):826–833

18. Dietzel M, Baltzer PAT, Dietzel A et al (2012) Artificial Neural Networks for differential diagnosis of breast lesions in MR-Mammography: a systematic approach addressing the influence of network architecture on diagnostic performance using a large clinical database. Eur J Radiol 81(7):1508–1513

19. Dietzel M, Baltzer PAT, Dietzel A et al (2010) Application of artificial neural networks for the prediction of lymph node metastases to the ipsilateral axilla - initial experience in 194 patients using magnetic resonance mammography. Acta Radiol 51(8):851–858

20. Dietzel M, Baltzer PAT, Vag T, Gajda M, Camara O, Kaiser WA (2010) The hook sign for differential diagnosis of malignant from benign lesions in magnetic resonance mammography: experience in a study of 1084 histologically verified cases. Acta Radiol 51(2):137–143

21. Dietzel M, Baltzer PAT, Vag T et al (2010) The adjacent vessel sign on breast MRI: new data and a subgroup analysis for 1,084 histologically verified cases. Korean J Radiol 11(2):178–186

22. Dietzel M, Baltzer PAT, Vag T et al (2010) The necrosis sign in magnetic resonancemammography: diagnostic accuracy in 1,084 histologically verified breast lesions. Breast J 16(6):603–608

23. Dietzel M, Baltzer PA, Vag T et al (2011) Magnetic resonance mammography in small vs. advanced breast lesions - systematic comparison reveals significant impact of lesion size on diagnostic accuracy in 936 histologically verified breast lesions. Rofo 183(2):126–135

24. Bland JM (2000) An introduction to medical statistics. Oxford University Press, Oxford

25. Sardanelli F, Boetes C, Borisch B et al (2010) Magnetic resonance imaging of the breast: recommendations from the EUSOMA working group. Eur J Cancer 46:1296–1316

26. Warren RML, Thompson D, Pointon LJ et al (2006) Evaluation of a prospective scoring system designed for a multicenter breast MR imaging screening study. Radiology 239:677–685

27. Gutierrez RL, DeMartini WB, Eby PR, Kurland BF, Peacock S, Lehman CD (2009) BI-RADS lesion characteristics predict likelihood of malignancy in breast MRI for masses but not for nonmasslike enhancement. AJR Am J Roentgenol 193:994–1000

28. Baltzer PAT, Benndorf M, Dietzel M, Gajda M, Runnebaum IB, Kaiser WA (2010) False-positive findings at contrast-enhanced breast MRI: a BI-RADS descriptor study. AJR Am J Roentgenol 194:1658–1663

29. Fischer DR, Wurdinger S, Boettcher J, Malich A, Kaiser WA (2005) Further signs in the evaluation of magnetic resonance mammography: a retrospective study. Invest Radiol 40:430–435

30. Sakamoto N, Tozaki M, Higa K et al (2008) Categorization of non-mass-like breast lesions detected by MRI. Breast Cancer 15:241–246