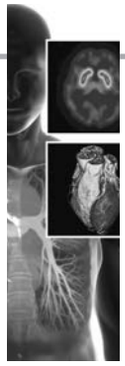

Computer-aided diagnostic models in breast cancer screening

Mammography is the most common modality for breast cancer detection and diagnosis and is often complemented by ultrasound and MRI. However, similarities between early signs of breast cancer and normal structures in these images make detection and diagnosis of breast cancer a difficult task. To aid physicians in detection and diagnosis, computer-aided detection and computer-aided diagnostic (CADx) models have been proposed. A large number of studies have been published for both computer-aided detection and CADx models in the last 20 years. The purpose of this article is to provide a comprehensive survey of the CADx models that have been proposed to aid in mammography, ultrasound and MRI interpretation. We summarize the noteworthy studies according to the screening modality they consider and describe the type of computer model, input data size, feature selection method, input feature type, reference standard and performance measures for each study. We also list the limitations of the existing CADx models and provide several possible future research directions.

KEYWORDS: breast cancer ▪ computer-aided detection ▪ computer-aided diagnosis ▪ mammography ▪ MRI ▪ ultrasound

Turgay Ayer[1], Mehmet US Ayvaci[1], Ze Xiu Liu[1], Oguzhan Alagoz[1,2] & Elizabeth S Burnside[†1,3]

[1]Industrial & Systems Engineering Department, University of Wisconsin, Madison, WI, USA
[2]Department of Population Health Sciences, University of Wisconsin, Madison, WI, USA
[3]Department of Biostatistics & Medical Informatics, University of Wisconsin, Madison, WI, USA
[†]Author for correspondence: Department of Radiology, University of Wisconsin Medical School, E3/311, 600 Highland Avenue, Madison, WI 53792-3252, USA
Tel.: +1 608 265 2021
Fax: +1 608 265 1836
eburnside@uwhealth.org

Radiological imaging, which often includes mammography, ultrasound (US) and MRI, is the most effective means, to date, for early detection of breast cancer [1]. However, differentiating between benign and malignant findings is difficult.

Successful breast cancer diagnosis requires systematic image analysis, characterization and integration of numerous clinical and mammographic variables [2], which is a difficult and error-prone task for physicians. This leads to low positive predictive value of imaging interpretation [3].

The integration of computer models into the radiological imaging interpretation process can increase the accuracy of image interpretation. There are two broad categories of computer models in breast cancer diagnosis: computer-aided detection (CADe) and computer-aided diagnostic (CADx) models. CADe models are computerized tools that assist radiologists in locating and identifying possible abnormalities in radiologic images, leaving the interpretation of the abnormality to the radiologist [4]. The potential for CADe models to improve detection of cancer has been investigated in several retrospective studies [5–8] as well as carefully controlled prospective studies [9–12]. For a review of CADe studies, the reader is referred to recent review articles by Hadjiiski *et al.* [13] and Nishikawa [14]. CADx models are decision aids for radiologists characterizing findings from radiologic images (e.g., size, contrast and shape) identified either by a radiologist or a CADe model [15]. CADx models have been demonstrated to increase the accuracy of mammography interpretation in several studies. Encouraged by promising results in mammography interpretation, numerous CADx models are being developed to help in breast US and MRI interpretation.

There are two reviews of CADx models, but neither are comprehensive in nature. The first, by Elter and Horsch, focuses on CADx models in mammography interpretation, but not in US and MRI, and concentrates on technical aspects of model development rather than more clinically relevant considerations [16]. The second, by Dorrius and van Ooijen, focuses on MRI CADx models [17]. Here we provide a comprehensive review for mammography, US and MRI CADx models in breast cancer diagnosis. We start by summarizing CADx models proposed for mammography interpretation. We then describe CADx models in US and MRI. We conclude by discussing several common limitations of existing research on CADx models and provide possible future research directions.

Mammography CADx models

Early work involving CADx models in mammography interpretation dates back to 1993. A summary list for primary mammography CADx models is presented in Table 1.

future medicine part of fsg

Table 1. Summary of computer-aided diagnostic models in mammography interpretation.

Study (year)	Size of dataset (n)	Model	AUC	Reader study	Ref.
Jiang *et al.* (1996)	107	ANN	0.92	Yes	[26]
Markopoulos *et al.* (2001)	240	ANN	0.937	Yes	[31]
Huo *et al.* (2002)	110	ANN	0.96	Yes	[35]
Floyd *et al.* (2000)	500	CBR	0.83	No	[37]
Elter *et al.* (2007)	2100	DT/CBR	0.87/0.89	No	[38]
Chan *et al.* (1999)	253	LDC	0.91	Yes	[34]
Gupta *et al.* (2006)	115	LDA	0.92	No	[41]
Wang *et al.* (1999)	419	BN	0.886	No	[42]
Chhatwal *et al.* (2009)	62,219	LR	0.963	Yes	[43]
Burnside *et al.* (2009)	62,219	BN	0.960	Yes	[44]
Ayer *et al.* (2010)	62,219	ANN	0.965	Yes	[45]
Bilska-Wolak *et al.* (2005)	151	LRbC	0.88	No	[46]

ANN: Artificial neural network; AUC: Area under the curve; BN: Bayesian network; CBR: Case-based reasoning; DT: Decision tree; LDA: Linear discriminant analysis; LDC: Linear discriminant classifier; LR: Logistic regression; LRbC: Likelihood ratio-based classifier.

Early work of CADx research used artificial neural networks (ANNs) and Bayesian networks (BNs). The first CADx model was proposed by Wu *et al.*, who developed an ANN to classify lesions detected by radiologists as malignant or benign [18]. They demonstrated that their simple ANN, which was built using 14 radiologist-extracted mammography features and trained on a small set of data, achieved higher area under the curve (AUC) of the receiver operating characteristic (ROC) curve than a group of attending radiologists without computer aid (0.89 vs 0.84). Baker *et al.* later built more complex ANN models, where the inputs included Breast Imaging Reporting and Data System (BI-RADS) descriptors as well as variables related to the patient's medical history [19]. Their approach was later extended and evaluated by others [20–23]. Fogel *et al.* also built one of the early ANN models that prospectively examined suspicious masses as a second opinion to radiologists [24]. Kahn *et al.* developed one of the first BN models to classify mammographic lesions as benign and malignant [25]. They used radiologist-extracted mammography features as the input to their model and demonstrated that BNs had a potential to help radiologists making diagnostic decisions.

Jiang *et al.* trained an ANN to differentiate malignant and benign clustered microcalcifications [26]. The microcalcifications were initially identified by the radiologists and eight features of these microcalcifications were automatically extracted by an image-processing algorithm. The training and testing data included 107 cases (40 malignant) from 53 patients. This retrospective study only included microcalcifications that underwent biopsy. Five radiologists participated in the observer study. ROC analysis was used to assess performance. The average cumulative AUC values for the ANN and the radiologists were 0.92 and 0.89, respectively. While the cumulative AUCs did not have a significant difference ($p = 0.22$), the comparison of AUCs over the 0.90 sensitivity threshold yielded statistically significant differences ($p < 0.05$). Jiang *et al.* later extended this model to classify lesions as malignant or benign for multiple-view mammograms [27]. They found that the use of a CADx model decreased the number of biopsied benign lesions while increasing the biopsy recommendations for malignant clusters. In a follow-up study, Jiang *et al.* demonstrated that, in addition to its diagnostic power, their ANN model had the potential to reduce the variability among radiologists in the interpretation of mammograms [28]. In another study, they compared their CADx model with independent double readings on 104 mammograms (46 malignant) containing clustered microcalcifications and reported more significant improvements in the ROC performance when the CADx model was used as compared with the independent double readings [29]. More recently, Rana *et al.* applied the CADx model developed by Jiang *et al.* on screen-film mammograms [26,27] to full-field digital mammograms [30]. They concluded that their CADx model maintained consistently high performance in classifying calcifications in full-field digital mammograms without requiring substantial modifications from its initial development on screen-film mammograms.

Markopoulos *et al.* compared three radiologists' diagnostic accuracies with or without computer aid [31]. The computer analysis utilized an ANN in diagnosis of clustered microcalcifications on mammograms. This retrospective study included 240 suspicious microcalcifications (108 malignant), which were identified by radiologists and extracted by an image-processing algorithm. The inputs to the ANN included eight features of the calcifications. Biopsy was the reference standard. The AUC of the CADx was 0.937, which was significantly higher than that of the physician with the highest performance (AUC = 0.835, $p = 0.012$). The authors concluded that CADx models also have the potential to help improve the diagnostic accuracy of radiologists.

Huo *et al.* also used ANNs to classify mass lesions detected on screen-film mammograms [32,33]. They automated the feature extraction process to reduce the intra-observer variability [28,34].

In a follow-up study, Huo *et al.* used different sets of data for training and testing instead of a single database [35]. Their database included 50 biopsy-proven malignant masses, 50 biopsy-proven benign masses and ten cysts proved by fine needle aspiration. The inputs to the ANN included four characteristics of masses (margin, sharpness, density and texture) that were automatically extracted by an image processing algorithm. When the CADx model was used, the average AUC of the radiologists increased from 0.93 to 0.96 ($p < 0.001$), demonstrating the generalizability of CADx models to distinct datasets. More recently, Li *et al.* converted the CADx model developed by Huo *et al.* on screen-film mammograms to apply to full-field digital mammograms [36]. They evaluated the performance of this CADx model using the AUC at various stages of the conversion process and concluded that CADx models had a potential to aid physicians in the clinical interpretation of full-field digital mammograms.

Floyd *et al.* proposed a case-based reasoning (CBR) approach, in which the classification is based on the ratio of the matched malignant cases to total matches in the database [37]. The primary advantage of the CBR method over an ANN is the transparent reasoning process that leads to the system's diagnosis. However, a key limitation of CBR is that a new case might not have any match in the database. This CBR analysis included 500 (174 malignant) cases. Of these 500 cases, 232 were masses alone, 192 were microcalcifications alone and 29 were combinations of masses and associated microcalcifications. The inputs to the CBR included ten features from the BI-RADS lexicon (five mass descriptors and five calcification descriptors) and a descriptor from clinical data. Biopsy was the reference standard. Two radiologists were asked to describe each lesion using the BI-RADS lexicon. The input dataset contained both retrospective (206 cases) and prospective (194 cases) data. The performance of the CBR model was compared with that of an ANN. While the ANN slightly outperformed the CBR (AUC = 0.86 vs 0.83, respectively), the study did not report statistical significance of this difference.

Elter *et al.* evaluated two novel CADx approaches that predicted breast biopsy outcomes [38]. The study retrospectively analyzed cases that contained masses or calcifications but not both. The dataset included 2100 masses (1045 malignant) and 1359 calcifications (610 malignant) that were extracted from mammograms in a public database and double reviewed by radiologists. The positive cases included histologically proven cancers, while negative cases were followed up for a 2-year period. The inputs to the CADx model included patient age and five features from the BI-RADS lexicon (two mass descriptors and three calcification descriptors). Elter *et al.* used two types of CADx systems: a decision tree and a CBR. An ANN was also implemented to compare its performance to that of the two proposed models. The models were evaluated based on ROC analysis. Contrary to the findings by Floyd *et al.* [37], they found that the CBR outperformed the ANN (AUC = 0.89 vs 88, respectively, $p < 0.001$), while the ANN performed better than the decision tree (AUC = 0.88 vs 0.87, respectively, $p < 0.001$). The authors concluded that both systems could potentially reduce the number of unnecessary biopsies with more accurate prediction of breast biopsy outcomes. However, the differences in AUC performances were small, raising the possibility that they may not be clinically significant.

Chan *et al.* retrospectively evaluated the effects of a linear discriminant classifier on radiologists' characterization of masses [34]. The dataset included 253 mammograms (127 malignant). Biopsy was the reference standard. The findings were initially identified by a radiologist and 41 features of these findings (texture and morphologic features) extracted by an image-processing algorithm were used as inputs to the linear discriminant classifier. Six reading radiologists evaluated the mammograms with and without CADx. The classification performance was evaluated by ROC analysis. The average AUC of the reading radiologists without CADx was 0.87 and improved to 0.91 with CADx ($p < 0.05$). Hadjiiski *et al.* performed similar studies to evaluate a CADx model and particularly investigated the extent of increase in diagnostic accuracy when more mammographic information was available [39,40]. Specifically, they evaluated two scenarios: the increase in the performance of CADx when trained on serial mammograms [39] and the increase in the performance of CADx when trained with interval change analysis, which used interval change information extracted from prior and current mammograms [40]. For both scenarios, they reported superior AUCs for the radiologists with CADx when compared with the radiologists without CADx (for the first scenario AUC = 0.85 vs 0.79, respectively, $p = 0.005$; and for the second scenario AUC = 0.87 vs 0.83, respectively, $p < 0.05$) and, thus, a significant improvement of the radiologists' diagnostic accuracy.

Gupta *et al.* retrospectively studied 115 biopsy-proven masses or calcification lesions (51 malignant) using a linear discriminant analysis (LDA)-based CADx model [41]. The images and case records were obtained from a public database. This study compared the performance of the LDA while using different descriptors for one mammographic view and two mammographic views. The attending radiologists described each abnormality using BI-RADS descriptors and categories. The inputs to the CADx model included patient age and two features from the BI-RADS lexicon (mass shape and mass margin). While the CADx with two mammographic views outperformed that with one mammographic view (AUC = 0.920 vs 0.881, respectively), the difference was not statistically significant ($p = 0.056$).

Wang *et al.* built and evaluated three BNs [42]. One of the BNs was constructed based on a total of 13 mammographic features and patients' characteristics. The other two BNs were hybrid classifiers, one of which was constructed by averaging the outputs from two subnetworks of mammographic-only or non-mammographic features. The third classifier used logistic regression (LR) to compute the outputs from the same subnetworks. This retrospective study included 419 cases (92 malignant). The verification of positive cases included biopsy and/or surgical reports, while negative cases were followed up for at least a 2-year period. The input features included four mammographic findings and nine descriptors from clinical data. The features were manually extracted by radiologists. The AUC for the BN that incorporated all 13 features was 0.886 and the AUCs for the BNs that included only mammographic features and patient characteristics were 0.813 and 0.713, respectively. The BN that included the full feature set was significantly better than both of the hybrid BNs ($p < 0.05$).

Recently, Chhatwal *et al.* [43] and Burnside *et al.* [44] developed a LR and BN, respectively, based on a consecutive dataset from a breast imaging practice consisting of 62,219 mammography records (510 malignant). The input features included 36 variables based on BI-RADS descriptors for masses, calcifications, breast density, associated findings and patients' clinical descriptors. The input dataset was recorded in the national mammography database format, which allowed the use of these models in other healthcare institutions. Contrary to most studies in the literature, they included the nonbiopsied mammograms in their training dataset and used cancer registries as the reference standard instead of the biopsy results. They analyzed the performance of the CADx models using ROC analysis and concluded that their CADx models performed better than that of the radiologists in aggregate (AUCs = 0.963 and 0.960 for LR and BN, respectively, vs 0.939 for the radiologist; $p < 0.05$). More recently, Ayer *et al.* developed an ANN model using the same dataset and demonstrated that the ANN model achieved slightly a higher AUC (0.965) than that of the LR and BN models as well as the radiologists [45]. Additionally, Ayer *et al.* extended the performance analysis of the CADx models from discrimination (classification) to calibration metrics, which assessed the ability of this ANN model to accurately predict the cancer risk for individual patients.

Bilska-Wolak *et al.* conducted a preclinical evaluation of a previously developed CADx model, a likelihood ratio-based classifier, on a new set of data [46]. The model retrospectively evaluated 151 new and independent cases (42 malignant). Biopsy was the reference standard. Suspicious masses were detected and described by an attending radiologist using 16 different features from the BI-RADS lexicon and patient history. The authors evaluated the CADx model based on ROC analysis and sensitivity statistics. The average AUC was 0.88. The model achieved 100% sensitivity at 26% specificity. The results were compared with an ANN model created using the same datasets. The AUC of the ANN was lower than that of the likelihood ratio-based classifier. Bilska-Wolak *et al.* concluded that their CADx model showed promising results that could reduce the number of false-positive mammograms.

US CADx models

Ultrasound imaging is an adjunct to diagnostic mammography, where CADx models could be used for improving diagnostic accuracy. CADx models developed for US scans date back to late 1990s. In this section, we review studies that apply CADx systems to breast sonography or US-mammography combination in distinguishing malignant from benign lesions. A summary list for the primary US CADx models is presented in Table 2.

Giger *et al.* classified malignant lesions in a database of 184 digitized US images [47]. Biopsy, cyst aspiration or image interpretation alone were used to confirm benign lesions, whereas malignancy was proven at biopsy. The authors utilized an LDA model to differentiate between benign

and malignant lesions using five computer-extracted features based on lesion shape and margin, texture, and posterior acoustic attenuation (two features). ROC analysis yielded AUCs of 0.94 for the entire database and 0.87 for the database that only included biopsy- and cyst-proven cases. The authors concluded that their analysis demonstrated that computerized analysis could improve the specificity of breast sonography.

Chen *et al.* developed an ANN to classify malignancies on US images [48]. A physician manually selected sub-images corresponding to a suspicious tumor region followed by computerized analysis of intensity variation and texture information. Texture correlation between neighboring pixels was used as the input to the ANN. The training and testing dataset included 140 biopsy-proven breast tumors (52 malignant). The performance was assessed by AUC, sensitivity and specificity metrics, which yielded an AUC of 0.956 with 98% sensitivity and 93% specificity at a threshold level of 0.2. The authors concluded that their CADx model was useful in distinguishing benign and malignant cases, yet also noted that larger datasets could be used to improve the performance.

Later, Chen *et al.* improved on a previous study [48] and devised an ANN model composed of three components: feature extraction, feature selection, and classification of benign and malignant lesions [49]. The study used two sets of biopsy-proven lesions; the first set with 160 digitally stored lesions (69 malignant) and the second set with 111 lesions (71 malignant) in hard-copy images that were obtained with the same US system. Hard-copy images were digitized using film scanners. Seven morphologic features were extracted from each lesion using an image-processing algorithm. Given the classifier, forward stepwise regression was employed to define the best performing features. These features were used as inputs to a two-layer feed-forward ANN. For the first set, the ANN achieved an AUC of 0.952, 90.6% sensitivity and 86.6% specificity. For the second set, the ANN achieved an AUC of 0.982, 96.7% sensitivity and 97.2% specificity. The ANN model trained on each dataset was demonstrated to be statistically extendible to other datasets at a 5% significance level. The authors concluded that their ANN model was an effective and robust approach for lesion classification, performing better than the counterparts published earlier [47,48].

Horsch *et al.* explored three aspects of an LDA classifier that was based on automatic segmentation of lesions and automatic extraction of lesion shape, margin, texture and posterior acoustic behavior [50]. The study was conducted using a database of 400 cases with 94 malignancies, 124 complex cysts and 182 benign lesions. The reference standard was either biopsy or aspiration. First, marginal benefit of adding a feature to the LDA model was investigated. Second, the performance of the LDA model in distinguishing carcinomas from different benign lesions was explored. The AUC values for the LDA model were 0.93 for distinguishing carcinomas from complex cysts and 0.72 for differentiating fibrocystic disease from carcinoma. Finally, eleven independent trials of training and testing were conducted to validate the LDA model. Validation resulted in a mean AUC of 0.87 when computer-extracted features from automatically delineated lesion margins were used. There was no statistically significant difference between the best two- and four-feature classifiers; therefore, adding features to the LDA model did not improve the performance.

Sahiner *et al.* investigated computer vision techniques to characterize breast tumors on 3D US volumetric images [51]. The dataset was composed of masses from 102 women who underwent either biopsy or fine-needle aspiration (56 had malignant masses). Automated mass segmentation in 2D and 3D, as well as feature extraction followed by LDA, were implemented to obtain malignancy scores. Stepwise feature selection was employed to reduce eight morphologic and 72 texture features into a best-feature subset. An AUC of 0.87 was achieved for the 2D-based classifier, while the AUC for the 3D-based classifier was 0.92. There was no statistically significant difference between the two classifiers ($p = 0.07$). The AUC values of the four radiologists fell in the range of 0.84 to 0.92. Comparing the performance of their model to that of radiologists, the difference was not statistically significant ($p = 0.05$). However, the

Table 2. Summary of computer-aided diagnostic models in ultrasound interpretation.

Study (year)	Size of dataset (n)	Model	AUC	Reader study	Ref.
Giger *et al.* (1999)	184	LDA	0.94	No	[47]
Chen *et al.* (1999)	140	ANN	0.956	No	[48]
Chen *et al.* (2003)	160/111	ANN	0.952/0.982	No	[49]
Horsch *et al* (2002)	400	LDA	0.87	No	[50]
Sahiner *et al.* (2004)	102	LDA	0.92	Yes	[51]
Drukker *et al.* (2008)	1046	BNN	0.90	Yes	[52]
Horsch *et al.* (2006)	717	BNN	0.91	Yes	[53]
Sahiner *et al.* (2009)	67	LDA	0.95	Yes	[54]

ANN: Artificial neural network; AUC: Area under the curve; BNN: Bayesian neural network; LDA: Linear discriminant analysis.

partial AUC for their model was significantly higher than those of the three radiologists ($p < 0.03$, 0.02 and 0.001).

Drukker *et al.* used various feature segmentation and extraction schemes as inputs to a Bayesian neural network (BNN) classifier with five hidden layers [52]. The purpose of the study was to evaluate a CADx workstation in a realistic setting representative of clinical diagnostic breast US practice. Benign or malignant lesions that were verified at biopsy or aspiration, as well as those determined through imaging characteristics on US scans, MR images and mammograms, were used for the analysis. The authors included nonbiopsied lesions in the dataset to make the series consecutive, which more accurately reflects clinical practice. The inputs to the network included lesion descriptors consisting of the depth:width ratio, radial gradient index, posterior acoustic signature and autocorrelation texture feature. The output of the network represented the probability of malignancy. The study was conducted on a patient population of 508 (101 had breast cancer) with 1046 distinct abnormalities (157 cancerous lesions). Comparing the current radiology practice with the CADx workstation, the CADx scheme achieved an AUC of 0.90, corresponding to 100% sensitivity at 30% specificity, while radiologists performed with 77% specificity for 100% sensitivity when only nonbiopsied lesions were included. When only biopsy-proven lesions were analyzed, computerized lesion characterization outperformed the radiologists.

In routine clinical practice, radiologists often combine the results from mammography and US, if available, when making diagnostic decisions. Several studies demonstrated that CADx could be useful in the differentiation of benign findings from malignant breast masses when sonographic data are combined with corresponding mammographic data. Horsch *et al.* evaluated and compared the performance of five radiologists with different expertise levels and five imaging fellows with or without the help of a BNN [53]. The BNN model utilized a computerized segmentation of the lesion. Mammographic features used as the input included spiculation, lesion shape, margin sharpness, texture and gray level. Sonographic input features included lesion shape, margin, texture and posterior acoustic behavior. All features were automatically extracted by an image-processing algorithm. This retrospective study examined a total of 359 (199 malignant) mammographic and 358 (67 malignant) sonographic images. Additionally, 97 (39 malignant) multimodality cases (both mammogram and sonogram) were used for testing purposes only. Biopsy was the reference standard. The performances of each radiologist/imaging fellow or pair of observers were quantified by the AUC, sensitivity and specificity metrics. Average AUC without BNN was 0.87 and with BNN was 0.92 ($p < 0.001$). The sensitivities without and with BNN were 0.88 and 0.93, respectively ($p = 0.005$). There was not a significant difference in specificities without and with BNN (0.66 vs 0.69, $p = 0.20$). The authors concluded that the performance of the radiologists and imaging fellows increased significantly with the help of the BNN model.

In another multimodality study, Sahiner *et al.* investigated the effect of a multimodal CADx system (using mammography and US data) in discriminating between benign and malignant lesions [54]. The dataset for the study consisted of 13 mammography features (nine morphologic, three spiculation and one texture) and eight 3D US features (two morphologic and six texture) that were extracted from 67 biopsy-proven masses (35 malignant). Ten experienced readers first gave a malignancy score based on mammography only, then re-evaluated based on mammography and US combined, and were finally allowed to change their minds given the CADx system's evaluation of the mass. The CADx system automatically extracted the features, which were then fed into a multimodality classifier (using LDA) to give a risk score. The results were compared using ROC curves, which suggested statistically significant improvement ($p = 0.05$) when the CADx system was consulted (average AUC = 0.95) over readers' assessment of combined mammography and US without the CADx (average AUC = 0.93). Sahiner *et al.* concluded that a CADx system combining the features from mammography and US may have the potential to improve radiologist's diagnostic decisions [54].

As discussed previously, a variety of sonographic features (texture, margin and shape) are used to classify benign and malignant lesions. 2D/3D Doppler imaging provides additional advantages in classification when compared with grayscale, by demonstrating breast lesion vascularity. Chang *et al.* extracted features of tumor vascularity from 3D power Doppler US images of 221 lesions (110 benign) and devised an ANN to classify lesions [55]. The study demonstrated that CADx, using 3D power Doppler imaging, can aid in the classification of benign and malignant lesions.

In addition to the aforementioned studies, there are other works that developed and evaluated CADx systems in differentiating between

benign and malignant lesions. Joo *et al.* developed an ANN that was demonstrated to have potential to increase the specificity of US characterization of breast lesions [56]. Song *et al.* compared an LR and an ANN in the context of differentiating between malignant and benign masses on breast sonograms from a small dataset [57]. There was no statistically significant difference between the performances of the two methods. Shen *et al.* investigated the statistical correlation between the computerized sonographic features, as defined by BI-RADS, and the signs of malignancy [58]. Chen and Hsiao evaluated US-based CADx systems by reviewing the methods used in classification [59]. They suggested the inclusion of pathologically specific tissue-and hormone-related features in future CADx systems. Gruszauskas *et al.* examined the effect of image selection on the performance of a breast US CADx system and concluded that their automated breast sonography classification scheme was reliable even with variation in user input [60]. Recently, Cui *et al.* published a study focusing on the development of an automated method segmenting and characterizing the breast masses on US images [61]. Their CADx system performed similarly whether it used automated segmentation or an experienced radiologist's segmentation. In a recent study, Yap *et al.* designed a survey to evaluate the benefits of computerized processing of US images in improving the readers' performance of breast cancer detection and classification [62]. The study demonstrated marginal improvements in classification when computer-processed US images alongside the originals are used in distinguishing benign from malignant lesions.

MRI CADx models

Dynamic contrast-enhanced MRI of the breast has been increasingly used in breast cancer evaluation and has been demonstrated to have potential to improve breast cancer diagnosis. The major advantage of MRI over other modalities is its ability to depict both morphologic and physiologic (kinetic enhancement) information [63]. Despite the advantages of MRI, it is a technology that is continuously evolving and is not currently cost effective for screening the general population [64,65]. Nevertheless, breast MRI is promising in terms of its high sensitivity, especially for high-risk young women with dense breasts. However, specificity has been highly variable in detection of breast cancer [17]. As a way of improving specificity, CADx models to aid discrimination of benign from malignant lesions in MRI imaging would be valuable. There are numerous CADx studies based on breast MRI. Generally, both morphologic and kinetic (enhancement) features are used in these studies to predict benign versus malignant breast lesions. In this section of the article, we only discuss the recent articles (published after 2003) that exemplify distinct aspects of breast MRI CADx research. A summary list for the primary MRI CADx models is presented in Table 3.

Szabó *et al.* used an ANN to retrospectively determine the discriminative ability of kinetic, morphologic and combined MRI features [66]. Inputs to the ANN included four morphologic and nine kinetic features from 105 biopsy-proven breast lesions with 75 malignancies. The model derived from the most relevant input variables, called the minimal model, resulted in the highest AUC value (0.771). The model with best kinetic features had an AUC of 0.743, the model with all features had an AUC of 0.727 and the model with qualitative architectural features, called the morphologic model, had an AUC of 0.678. The expert radiologists achieved an AUC of 0.799; therefore, the performance was comparable to that of the minimized model.

Nattkemper *et al.* analyzed various machine learning methods using four morphologic and five kinetic tumor features found on MRI as input [67]. The investigated methods included k-means clustering, self-organizing maps, Fisher discriminant analysis, k-nearest classifiers, support vector machines and decision tree. The input dataset included dynamic contrast-enhanced MRI data of 74 breast lesions (49 malignant). Biopsy was the reference standard. Among the investigated methods, support vector machines achieved the highest AUC (0.88). They also demonstrated that, among all the MRI features analyzed, the wash-out type features extracted by radiologists improved classification performance the most.

Meinel *et al.* developed an MRI CADx system to improve radiologists' performance in classifying breast lesions [68]. An ANN was constructed using 80 biopsy-proven lesions (43 malignant). Inputs to the ANN were the best 13 features

Table 3. Summary of computer-aided diagnostic models in MRI interpretation.

Study (year)	Size of dataset (n)	Model	AUC	Reader study	Ref.
Szabó *et al.* (2004)	105	ANN	0.771	Yes	[66]
Nattkemper *et al.* (2005)	74	SVM	0.88	No	[67]
Meinel *et al.* (2007)	80	ANN	0.907	Yes	[68]
Deurloo *et al.* (2005)	100	LR	0.91	Yes	[69]

ANN: Artificial neural network; AUC: Area under the curve; LR: Logistic regression; SVM: Support vector machine.

from a set of 42, based on lesion shape, texture and enhancement kinetics information. The performance was assessed by comparison of AUC values from five human readers diagnosing the tumor with and without the help of the CADx system. When only the first abnormality shown to human readers was included, ROC analysis yielded AUCs of 0.907 with ANN assistance and 0.816 without the assistance. The difference was statistically significant ($p < 0.011$); therefore, Meinel *et al.* demonstrated that their ANN model improves the performance of human readers.

Deurloo *et al.* combined the clinical assessment of clinically and mammographically occult breast lesions by radiologists with computer-calculated probability of malignancy of each lesion into an LR model [69]. Inputs to the LR model included the four best features from a set of six morphologic and three temporal features. Either biopsy-proven lesions or lesions showing transient enhancement were included in the study. The difference between the performance of clinical readings (AUC = 0.86) and computerized analysis (AUC = 0.85) was not statistically significant ($p = 0.99$). However, the combined model performed significantly higher (AUC = 0.91, $p = 0.03$) when compared with clinical reading without computerized analysis. The results demonstrated how computerized analysis could complement clinical interpretation of magnetic resonance images.

There are several other studies that addressed the use of CADx systems in MRI of the breast. Williams *et al.* evaluated the sensitivity of computer-generated kinetic features from CADstream, the first CADx system for breast MRI, for 154 biopsy-proven lesions (41 malignant) [70]. The study suggested that computer-aided classification improved radiologists' performance. Lehman *et al.* compared the accuracy of breast MRI assessments with and without the same software, CADstream [71]. They concluded that the software may improve the accuracy of radiologists' interpretation; however, the study was conducted on a small set of 33 lesions (nine malignant). Nie *et al.* investigated the feasibility of quantitative analysis of MRI images [72]. Morphology/texture features of breast lesions were selected by an ANN and used in the classification of benign and malignant lesions. Baltzer *et al.* investigated the incremental diagnostic value of complete enhancing lesions using a CADx model [73]. The study reported improvement in specificity with no statistical significance. In a different study, Baltzer *et al.* investigated both automated and manual measurement methods to assess contrast enhancement kinetics [74]. They analyzed and compared evaluation of contrast enhancements via curve-type assessment by radiologists, region of interest and CADx. The methods proved diagnostically useful although no statistically significant difference was found.

Future perspective

There have been significant advances in CADx models in the last 20 years. However, several issues remain open for future researchers. First and most notably, almost all of the existing CADx models are trained and tested on retrospectively collected cases that may not represent the real clinical practice. Large prospective studies are required to evaluate the performance of CADx models in real life before employing them in a clinical setting.

Second, an objective comparative performance evaluation of the existing CADx models is difficult because the reported performances depend on the dataset used in model building. One approach to a systematic performance comparison would be to use large and consistent, publicly available datasets for testing purposes. However, although this approach will give some idea about the realistic/comparable performances of the CADx systems, it would not be completely accurate because a CADx model performing the best on one dataset might be outperformed by another CADx model on another dataset.

Third, a frequently ignored issue in CADx model development is the clinical interpretability of the model. Aspects of the CADx model that allow clinical interpretations significantly influence the acceptance of the CADx model by the physicians. Most of the existing CADx models are based on ANNs. Although ANNs are powerful in terms of their predictive abilities, their parameters do not carry any real-life interpretation, hence, they are often referred to as 'black boxes'. Other models such as LR, BN or CBR allow direct clinical interpretation. However, the number of such studies is significantly limited as compared with ANN models.

Fourth, performance assessment of the CADx models are usually limited to discrimination (classification) metrics (e.g., sensitivity, specificity and AUC). On the other hand, the accuracy of risk prediction for individual patients, referred to as calibration, is often ignored. Although discrimination assesses the ability to correctly distinguish between benign and malignant abnormalities, it does not tell much about the accuracy of risk prediction for individual patients [75]. However, clinical decision-making usually involves decisions for individual patients under uncertainty; therefore,

it is aided more effectively by accurate risk estimates [75,76]. That is, calibration is sometimes as equivalently important as discrimination; therefore, future studies should consider measuring calibration performance as well.

Last but not least, breast cancer diagnosis often involves information collected from several sources, such as information from multiple mammographic views, prior screening history and additional examinations (e.g., US and MRI). However, most CADx models are built to process the information obtained from a single source. Therefore, future CADx models should aim to incorporate all possible information from various sources when making recommendations to radiologists.

Financial & competing interests disclosure

E Burnside is funded by NIH grants R01CA127379 and K07CA114181, and O Alagoz is funded by NIH grants R21CA129393 and 1UL1RR025011 from the CTSA program of NCRR NIH. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Executive summary

Computer-aided detection & diagnosis

- Although radiological imaging is the most effective means of early detection of breast cancer, differentiating between benign and malignant findings is difficult.
- To aid physicians in detection and diagnosis, computer-aided detection and computer-aided diagnostic (CADx) models have been proposed.
- Computer-aided detection models are computerized models that assist radiologists in locating and identifying possible abnormalities in radiologic images.
- CADx models are decision aids to radiologists in characterizing findings from radiologic images identified either by a radiologist or a computer-aided detection model.
- In this article, we provide a comprehensive review of CADx models developed for radiologic imaging modalities (i.e., mammography, ultrasound and MRI) in breast cancer diagnosis.

Future perspective

- Large prospective studies are required to evaluate the performance of CADx models in real life before they can be used in clinics.
- CADx models that allow direct clinical interpretation are encouraged.
- The use of large and consistent publicly available datasets is recommended for testing and comparing various available CADx models.
- Performance evaluation of CADx models should not be limited to discrimination, but should be extended to calibration.
- Future CADx models should aim to incorporate all possible information from different sources when making recommendations to radiologists.

Bibliography

Papers of special note have been highlighted as:
▪ of interest

1 Reddy M, Given-Wilson R: Screening for breast cancer. *Womens Health Med.* 3(1), 22–27 (2006).

2 Giger ML: Computer-aided diagnosis in radiology. *Acad. Radiol.* 9(1), 18–25 (2002).

3 Malich A, Boehm T, Facius M *et al.*: Differentiation of mammographically suspicious lesions: evaluation of breast ultrasound, MRI mammography and electrical impedance scanning as adjunctive technologies in breast cancer detection. *Clin. Radiol.* 56(4), 278–283 (2001).

4 Burhenne L, Wood S, D'Orsi C *et al.*: Potential contribution of computer-aided detection to the sensitivity of screening mammography. *Radiology* 215(2), 554–562 (2000).

5 Birdwell R, Ikeda D, O'Shaughnessy K, Sickles E: Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection. *Radiology* 219(1), 192–202 (2001).

6 Brem RF, Baum J, Lechner M *et al.*: Improvement in sensitivity of screening mammography with computer-aided detection: a multiinstitutional trial. *AJR Am. J. Roentgenol.* 181(3), 687–693 (2003).

7 Malich A, Sauner D, Marx C *et al.*: Influence of breast lesion size and histologic findings on tumor detection rate of a computer-aided detection system. *Radiology* 228(3), 851–856 (2003).

8 Malich A, Schmidt S, Fischer DR, Facius M, Kaiser WA: The performance of computer-aided detection when analyzing prior mammograms of newly detected breast cancers with special focus on the time interval from initial imaging to detection. *Eur. J. Radiol.* 69(3), 574–578 (2009).

9 Birdwell R, Bandodkar P, Ikeda D: Computer-aided detection with screening mammography in a university hospital setting. *Radiology* 236(2), 451–457 (2005).

10 Cupples TE, Cunningham JE, Reynolds JC: Impact of computer-aided detection in a regional screening mammography program. *AJR Am. J. Roentgenol.* 185(4), 944–950 (2005).

11 Dean JC, Ilvento CC: Improved cancer detection using computer-aided detection with diagnostic and screening mammography: prospective study of 104 cancers. *AJR Am. J. Roentgenol.* 187(1), 20–28 (2006).

12 Freer T, Ulissey M: Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology* 220(3), 781–786 (2001).

13 Hadjiiski L, Sahiner B, Chan HP: Advances in computer-aided diagnosis for breast cancer. *Curr. Opin. Obstet. Gynecol.* 18(1), 64–70 (2006).

14 Nishikawa RM: Current status and future directions of computer-aided diagnosis in mammography. *Comput. Med. Imaging Graph.* 31(4), 224–235 (2007).

15 Vyborny CJ, Giger ML, Nishikawa RM: Computer-aided detection and diagnosis of breast cancer. *Radiol. Clin. North Am.* 38(4), 725–740 (2000).

16 Elter M, Horsch A: CADx of mammographic masses and clustered microcalcifications: a review. *Med. Phys.* 36(6), 2052–2068 (2009).

17 Dorrius MD, van Ooijen PMA: Computer-aided detection in breast magnetic resonance imaging: a review. *Imaging Decisions* 12(3), 29–36 (2008).

18 Wu Y, Giger M, Doi K, Vyborny C, Schmidt R, Metz C: Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. *Radiology* 187(1), 81–87 (1993).

19 Baker J, Kornguth P, Lo J, Williford M, Floyd C Jr: Breast cancer: prediction with artificial neural network based on BI-RADS standardized lexicon. *Radiology* 196(3), 817–822 (1995).

▪ **One of the earliest mammography computer-aided diagnostic (CADx) models built using artificial neural networks.**

20 Lo JY, Baker JA, Kornguth PJ, Floyd CE Jr: Effect of patient history data on the prediction of breast cancer from mammographic findings with artificial neural networks. *Acad. Radiol.* 6(1), 10–15 (1999).

21 Lo JY, Baker JA, Kornguth PJ, Iglehart JD, Floyd CE Jr: Predicting breast cancer invasion with artificial neural networks on the basis of mammographic features. *Radiology* 203(1), 159–163 (1997).

22 Lo JY, Markey MK, Baker JA, Floyd CE Jr: Cross-institutional evaluation of BI-RADS predictive model for mammographic diagnosis of breast cancer. *AJR Am. J. Roentgenol.* 178(2), 457–463 (2002).

23 Markey MK, Lo JY, Floyd CE Jr: Differences between computer-aided diagnosis of breast masses and that of calcifications 1. *Radiology* 223(2), 489–493 (2002).

24 Fogel DB, Wasson EC, Boughton EM, Porto VW: A step toward computer-assisted mammography using evolutionary programming and neural networks. *Cancer Lett.* 119(1), 93–97 (1997).

25 Kahn CE, Roberts LM, Shaffer KA, Haddawy P: Construction of a Bayesian network for mammographic diagnosis of breast cancer. *Comput. Biol. Med.* 27(1), 19–29 (1997).

26 Jiang Y, Nishikawa RM, Wolverton DE *et al.*: Malignant and benign clustered microcalcifications: automated feature analysis and classification. *Radiology* 198(3), 671–678 (1996).

27 Jiang Y, Nishikawa RM, Schmidt RA, Metz CE, Giger ML, Doi K: Improving breast cancer diagnosis with computer-aided diagnosis. *Acad. Radiol.* 6(1), 22–33 (1999).

28 Jiang Y, Nishikawa RM, Schmidt RA, Toledano AY, Doi K: Potential of computer-aided diagnosis to reduce variability in radiologists' interpretations of mammograms depicting microcalcifications. *Radiology* 220(3), 787–794 (2001).

29 Jiang Y, Nishikawa RM, Schmidt RA, Metz CE: Comparison of independent double readings and computer-aided diagnosis (CAD) for the diagnosis of breast calcifications. *Acad. Radiol.* 13(1), 84–94 (2006).

30 Rana RS, Jiang Y, Schmidt RA, Nishikawa RM, Liu B: Independent evaluation of computer classification of malignant and benign calcifications in full-field digital mammograms. *Acad. Radiol.* 14(3), 363–370 (2007).

31 Markopoulos C, Kouskos E, Koufopoulos K, Kyriakou V, Gogas J: Use of artificial neural networks (computer analysis) in the diagnosis of microcalcifications on mammography. *Eur. J. Radiol.* 39(1), 60–65 (2001).

32 Huo Z, Giger ML, Vyborny CJ, Wolverton DE, Metz CE: Computerized classification of benign and malignant masses on digitized mammograms: a study of robustness. *Acad. Radiol.* 7(12), 1077–1084 (2000).

33 Huo Z, Giger ML, Vyborny CJ, Wolverton DE, Schmidt RA, Doi K: Automated computerized classification of malignant and benign masses on digitized mammograms. *Acad. Radiol.* 5(3), 155–168 (1998).

34 Chan HP, Sahiner B, Helvie MA *et al.*: Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis: an ROC study. *Radiology* 212(3), 817–827 (1999).

35 Huo Z, Giger ML, Vyborny CJ, Metz CE: Breast cancer: effectiveness of computer-aided diagnosis observer study with independent database of mammograms. *Radiology* 224(2), 560–568 (2002).

36 Li H, Giger ML, Yuan Y *et al.*: Evaluation of computer-aided diagnosis on a large clinical full-field digital mammographic dataset. *Acad. Radiol.* 15(11), 1437–1445 (2008).

37 Floyd CE Jr, Lo JY, Tourassi GD: Case-based reasoning computer algorithm that uses mammographic findings for breast biopsy decisions. *AJR Am. J. Roentgenol.* 175(5), 1347–1352 (2000).

38 Elter M, Schulz-Wendtland R, Wittenberg T: The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process. *Med. Phys.* 34(11), 4164–4172 (2007).

39 Hadjiiski L, Chan HP, Sahiner B *et al.*: Improvement in radiologists' characterization of malignant and benign breast masses on serial mammograms with computer-aided diagnosis: an ROC study. *Radiology* 233(1), 255–265 (2004).

40 Hadjiiski L, Sahiner B, Helvie MA *et al.*: Breast masses: computer-aided diagnosis with serial mammograms. *Radiology* 240(2), 343–356 (2006).

41 Gupta S, Chyn PF, Markey MK: Breast cancer CAD based on BI-RADS descriptors from two mammographic views. *Med. Phys.* 33(6), 1810–1817 (2006).

42 Wang XH, Zheng B, Good WF, King JL, Chang YH: Computer-assisted diagnosis of breast cancer using a data-driven Bayesian belief network. *Int. J. Med. Inform.* 54(2), 115–126 (1999).

43 Chhatwal J, Alagoz O, Lindstrom MJ, Kahn CE, Shaffer KA, Burnside ES: A logistic regression model based on the national mammography database format to aid breast cancer diagnosis. *AJR Am. J. Roentgenol.* 192(4), 1117–1127 (2009).

44 Burnside ES, Davis J, Chhatwal J *et al.*: Probabilistic computer model developed from clinical data in national mammography database format to classify mammographic findings. *Radiology* 251(3), 663–672 (2009).

45 Ayer T, Alagoz O, Chhatwal J, Shavlik JW, Kahn CE, Burnside ES: An artificial neural network to quantify malignancy risk based on mammography findings: discrimination and calibration. *Cancer* (2010) (In Press).

▪ **Unlike most of the CADx models in the literature, this study considers not only the ability of the CADx model to differentiate between benign and malignant findings, but also its ability to accurately estimate breast cancer risk for individual patients, which might be especially valuable in stratifying patients into lower and higher risk groups and in risk communication.**

46 Bilska-Wolak AO, Floyd CE, Lo JY, Baker JA: Computer aid for decision to biopsy breast masses on mammography validation on new cases 1. *Acad. Radiol.* 12(6), 671–680 (2005).

47 Giger M, Al-Hallaq H, Huo Z *et al.*: Computerized analysis of lesions in US images of the breast. *Acad. Radiol.* 6(11), 665–674 (1999).

▪ **One of the earliest ultrasound CADx models.**

48 Chen D, Chang R, Huang Y: Computer-aided diagnosis applied to US of solid breast nodules by using neural networks 1. *Radiology* 213(2), 407–412 (1999).

49 Chen CM, Chou YH, Han KC *et al.*: Breast lesions on sonograms: computer-aided diagnosis with nearly setting-independent features and artificial neural networks. *Radiology* 226(2), 504–514 (2003).

50 Horsch K, Giger M, Venta L, Vyborny C: Computerized diagnosis of breast lesions on ultrasound. *Med. Phys.* 29(1), 157–164 (2002).

51 Sahiner B, Chan HP, Roubidoux MA *et al.*: Computerized characterization of breast masses on three-dimensional ultrasound volumes. *Med. Phys.* 31(4), 744–754 (2004).

52 Drukker K, Gruszauskas NP, Sennett CA, Giger ML: Breast US computer-aided diagnosis workstation: performance with a large clinical diagnostic population. *Radiology* 248(2), 392–397 (2008).

53 Horsch K, Giger ML, Vyborny CJ, Lan L, Mendelson EB, Hendrick RE: Classification of breast lesions with multimodality computer-aided diagnosis: observer study results on an independent clinical dataset. *Radiology* 240(2), 357–368 (2006).

54 Sahiner B, Chan H-P, Hadjiiski LM *et al.*: Multi-modality CADx: ROC study of the effect on radiologists' accuracy in characterizing breast masses on mammograms and 3D ultrasound images. *Acad. Radiol.* 16(7), 810–818 (2009).

55 Chang RF, Huang SF, Moon WK, Lee YH, Chen DR: Solid breast masses: neural network analysis of vascular features at three-dimensional power Doppler US for benign or malignant classification. *Radiology* 243(1), 56–62 (2007).

56 Joo S, Yang Y, Moon W, Kim H: Computer-aided diagnosis of solid breast nodules: use of an artificial neural network based on multiple sonographic features. *IEEE Transactions on Medical Imaging* 23(10), 1292–1300 (2004).

57 Song JH, Venkatesh SS, Conant EA, Arger PH, Sehgal CM: Comparative analysis of logistic regression and artificial neural network for computer-aided diagnosis of breast masses. *Acad. Radiol.* 12(4), 487–495 (2005).

58 Shen WC, Chang RF, Moon WK, Chou YH, Huang CS: Breast ultrasound computer-aided diagnosis using BI-RADS features. *Acad. Radiol.* 14(8), 928–939 (2007).

59 Chen DR, Hsiao YH: Computer-aided diagnosis in breast ultrasound. *J. Med. Ultrasound* 16(1), 46–56 (2008).

60 Gruszauskas NP, Drukker K, Giger ML, Sennett CA, Pesce LL: Performance of breast ultrasound computer-aided diagnosis: dependence on image selection. *Acad. Radiol.* 15(10), 1234–1245 (2008).

61 Cui J, Sahiner B, Chan HP *et al.*: A new automated method for the segmentation and characterization of breast masses on ultrasound images. *Med. Phys.* 36(5), 1553–1565 (2009).

62 Yap MH, Edirisinghe E, Bez H: Processed images in human perception: a case study in ultrasound breast imaging. *Eur. J. Radiol.* 73(3), 682–687 (2010).

63 Suri JS, Rangayyan RM (Eds): *Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer*. The International Society for Optical Engineering, WA, USA (2006).

64 Moore SG, Shenoy PG, Fanucchi L, Tumeh JW, Flowers CR: Cost-effectiveness of MRI compared with mammography for breast cancer screening in a high risk population. *BMC Health Serv. Res.* 9, 9 (2009).

65 Plevritis SK, Kurian AW, Sigal BM *et al.*: Cost-effectiveness of screening *BRCA1/2* mutation carriers with breast magnetic resonance imaging. *JAMA* 295(20), 2374–2384 (2006).

66 Szabó BK, Wiberg MK, Bone B, Aspelin P: Application of artificial neural networks to the analysis of dynamic MR imaging features of the breast. *Eur. J. Radiol.* 14(7), 1217–1225 (2004).

▪ **One of the earliest MRI CADx models.**

67 Nattkemper T, Arnrich B, Lichte O *et al.*: Evaluation of radiological features for breast tumour classification in clinical screening with machine learning methods. *Artif. Intell. Med.* 34(2), 129–139 (2005).

68 Meinel LA, Stolpen AH, Berbaum KS, Fajardo LL, Reinhardt JM: Breast MRI lesion classification: improved performance of human readers with a backpropagation neural network computer-aided diagnosis (CAD) system. *J. Magn. Reson. Imaging* 25(1), 89–95 (2007).

69 Deurloo E, Muller S, Peterse J, Besnard A, Gilhuijs K: Clinically and mammographically occult breast lesions on MR images: potential effect of computerized assessment on clinical reading 1. *Radiology* 234(3), 693–701 (2005).

70 Williams T, DeMartini W, Partridge S, Peacock S, Lehman C: Breast MR imaging: computer-aided evaluation program for discriminating benign from malignant Lesions 1. *Radiology* 244(1), 94–103 (2007).

71 Lehman C, Peacock S, DeMartini W, Chen X: A new automated software system to evaluate breast MR examinations: improved specificity without decreased sensitivity. *AJR Am. J. Roentgenol.* 187(1), 51–56 (2006).

72 Nie K, Chen JH, Yu HJ, Chu Y, Nalcioglu O, Su MY: Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI. *Acad. Radiol.* 15(12), 1513–1525 (2008).

73 Baltzer PAT, Renz DM, Kullnig PE, Gajda M, Camara O, Kaiser WA: Application of computer-aided diagnosis (CAD) in MR-mammography (MRM): do we really need whole lesion time curve distribution analysis? *Acad. Radiol.* 16(4), 435–442 (2009).

74 Baltzer PAT, Freiberg C, Beger S *et al.*: Clinical MR-mammography: are computer-assisted methods superior to visual or manual measurements for curve type analysis? A Systematic Approach. *Acad. Radiol.* 16(9), 1070–1076 (2009).

75 Cook NR: Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 115(7), 928–935 (2007).

76 Cook NR: Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin. Chem.* 54(1), 17–23 (2008).