

# Pattern Recognition Approaches for Breast Cancer DCE-MRI Classification: A Systematic Review

Roberta Fusco<sup>1,2</sup> · Mario Sansone<sup>2</sup> · Salvatore Filice<sup>1</sup> · Guglielmo Carone<sup>1</sup> · Daniela Maria Amato<sup>1</sup> · Carlo Sansone<sup>2</sup> · Antonella Petrillo<sup>1</sup>

Received: 30 September 2015 / Accepted: 29 March 2016 / Published online: 31 August 2016  
© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** We performed a systematic review of several pattern analysis approaches for classifying breast lesions using dynamic, morphological, and textural features in dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI). Several machine learning approaches, namely artificial neural networks (ANN), support vector machines (SVM), linear discriminant analysis (LDA), tree-based classifiers (TC), and Bayesian classifiers (BC), and features used for classification are described. The findings of a systematic review of 26 studies are presented. The sensitivity and specificity are respectively 91 and 83 % for ANN, 85 and 82 % for SVM, 96 and 85 % for LDA, 92 and 87 % for TC, and 82 and 85 % for BC. The sensitivity and specificity are respectively 82 and 74 % for dynamic features, 93 and 60 % for morphological features, 88 and 81 % for textural features, 95 and 86 % for a combination of dynamic and morphological features, and 88 and 84 % for a combination of dynamic, morphological, and other features. LDA and TC have the best performance. A combination of dynamic and morphological features gives the best performance.

**Keywords** Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) · Breast cancer · Pattern recognition approach · Classification

## 1 Introduction

Breast cancer is the most common cancer among women in the Western world. It is the second leading cause of cancer death in women today (after lung cancer) and is estimated to cause 15 % of cancer deaths [1]. Therefore, screening for early diagnosis of breast cancer is of great interest.

The currently widespread screening method is RX mammography, which plays an important role in clinical practice [2, 3]. However, this method has some drawbacks: it uses ionizing radiation, it is not adequate for young women because of their high-density breasts, and detection of breast lesions is difficult because of the lack of functional information. Breast ultrasound (US) is able to detect additional cancers in women with dense breasts and negative mammography and is helpful for the characterization of mammographically detected abnormalities, evaluation of tumor size and nodal status, and guiding needle biopsy [4]. However, it is of limited value in detecting additional ipsi- or contra-lateral malignant lesions.

Magnetic resonance imaging (MRI) and in particular the emerging methodology of dynamic contrast-enhanced (DCE)-MRI has demonstrated great potential in the screening of high-risk women, staging newly diagnosed breast cancer patients, and assessing therapy effects thanks to its minimal invasiveness and ability to visualize dynamic (functional) information not available with conventional imaging. Therefore MRI, and in particular DCE-MRI, is gaining popularity as an important complementary diagnostic tool for the early detection of breast cancer [2–5].

Roberta Fusco and Mario Sansone have contributed equally to the manuscript.

✉ Roberta Fusco  
r.fusco@istitutotumori.na.it

<sup>1</sup> Department of Diagnostic Imaging, metabolic and radiant Therapy, National Cancer Institute of Naples “Pascale Foundation”, Via Mariano Semmola 80131, Naples, Italy

<sup>2</sup> Department of Electrical Engineering and Information Technologies, University ‘Federico II’, Via Claudio 80125, Naples, Italy

MRI is currently used as a complement to conventional X-ray mammography in the diagnosis of breast lesions [2]. It has been shown that 17–34 % of cancer foci visible on breast MRI are not detected by mammography. Because of the higher cost and increased time required to read an MRI data set ( $\sim 400$  images per patient), MRI will probably never be a complete replacement for mammography, but it is certainly an excellent screening tool for high-risk patients. Reducing the workload required to read an MRI data set would make it a more practical clinical screening tool. Therefore, the development of methods using low-cost hardware for lesion detection and classification is of great interest. X-ray mammography remains the gold standard for breast cancer screening and offers high two-dimensional (2D) resolution, which is advantageous for detecting small variations in tissue composition, such as micro-calcifications [3].

However, due to the constraints of imaging a three-dimensional (3D) structure in a single plane, breast US or DCE-MRI is often used as a secondary imaging technique when a suspicious lesion is found using mammography [3, 5]. DCE-MRI is also very good at imaging dense breasts, but its major advantages over mammography and US are the ability to (a) image the entire breast as thin slices that comprise the entire breast volume and (b) measure variations in contrast uptake that provide information about the vascularity of the breast tissue [6].

On account of breast DCE-MRI's high 3D resolution and its ability to acquire kinetic contrast information, its lesion detection sensitivity is close to 100 % [7], much higher than that of either mammography or US [1]. However, the specificity of breast DCE-MRI is low, with reported rates of between 30 and 70 % [7, 8].

In addition to the problem of low specificity, another shortcoming of breast MRI is that only experienced radiologists are able to accurately distinguish benign from malignant tumors [1, 9]. This often leads to high rates of inter-observer variability [9]. Therefore, one of the challenges in facilitating increased acceptance of breast DCE-MRI as a screening modality is reducing false positive detection errors, thereby boosting detection specificity. Additionally, the inter-observer variability for breast DCE-MRI must be minimized.

For these reasons, several authors have proposed using various features in DCE-MRI images to decide whether a given tumor is benign or malignant. For example, radiologists differentiate tumors based on features that describe the biological activity of the tumor using dynamic parameters (vascularization, permeability, flux) [10–22], tumor size, tumor boundary shape (morphological characteristics), or tumor heterogeneity (textural features) [23–49].

Computer-aided diagnosis (CAD) systems, using pattern analysis approaches, have the potential to assist radiologists in the detection and classification of breast cancer. A key component of the development of such CAD systems is the selection of an appropriate classification function responsible for separating malignant and benign lesions.

In the last two decades, many studies have addressed the problem of tumor lesion classification based on DCE-MRI data analysis. It has been recognised that this problem can be addressed in a pattern recognition framework with the use of opportune features and classifiers.

Despite large effort, there is still no agreement on the features most suitable for this task. Many kind of features have been used. Dynamic features take into account the time course of the contrast agent within the lesion, but they can fail to describe other features of the lesion, such as tumor heterogeneity. Textural features have thus been introduced. Compartmental modeling can add useful information concerning vascular permeability. Furthermore, morphological features have been traditionally used in tumor classification and they can be added to other dynamic features, with the additional advantage that tumor morphology can be delineated more precisely using the dynamic information of DCE-MRI. Spatiotemporal features have been suggested to combine spatial and dynamic information.

Similarly, it is not clear what kind of classifiers can give the best performance. Several classifiers have been used. It is generally recognised that tree-based classifiers are more easily accepted by humans because they can require to simple threshold-based rules; however, they can suffer from over-fitting. Linear classifiers are also easy to design and understand, but linear combinations of features do not always have simple interpretations. More sophisticated classifiers such as artificial neural networks (ANN) and support vector machine (SVM) are strongly non-linear and thus classification hyperplanes are difficult to interpret. Moreover, not all classifiers work well with all types of features; therefore, various combinations of classifier-features have been attempted.

This study surveys the literature of the last two decades that focuses on features and classifiers used for the classification of tumor lesions detected from DCE-MRI data. We performed a systematic review of several machine learning algorithms proposed in the literature for classifying breast lesions using dynamic [10–22], morphological, and textural features in DCE-MRI [23–49]. This study considers the following machine learning approaches: ANN, SVM, linear discriminant analysis (LDA), tree-based classifiers (TC), and Bayesian classifiers (BC). This systematic review is conducted using a meta-analysis. As such, our objective is not to present a short summary of all studies, but instead to focus on aspects common to all

studies and present a statistical analysis of the performance of the algorithms in the literature.

## 2 Review of Methodology

### 2.1 Search Criteria

Several electronic databases were searched, namely PubMed (US National Library of Medicine, <http://www.ncbi.nlm.nih.gov/pubmed>), Scopus (Elsevier, <http://www.scopus.com/>), Web of Science (Thomson Reuters, <http://apps.webofknowledge.com/>), and Google Scholar (<https://scholar.google.it/>). The following search criteria were used: “breast cancer” and “breast lesions” for the clinical domain and “DCE-MRI”, “Dynamic Contrast Enhanced-MRI”, and “Dynamic Contrast Enhanced-Magnetic Resonance Imaging” for the diagnostic test. To make sure that no study was missed, a free-text search was also performed. The search covered the years from 1995 through 2014. Furthermore, all reference lists of the obtained papers were scrutinized for studies not indexed in the electronic databases.

If not otherwise stated, all the studies reviewed herein fulfill the following criteria: (1) thorough clinical characterization of the patients with DCE-MRI (studies using other diagnostic techniques were excluded); (2) specification of applied classifiers; (3) accuracy of classifier reported in terms of sensitivity and specificity; and (4) used one of the following classifiers: ANN, SVM, LDA, TC, or BC.

In the present review, all relevant studies were scrutinized, but only studies that satisfied the inclusion criteria are included in the review (Fig. 1). Furthermore, this analysis was carried out only for studies of subjects with breast lesions. Information extracted from each study included the title, authors, year of publication, sample size, age of subjects, reference standard, and numbers of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).

### 2.2 Pattern Recognition Approach

As many textbooks are available on this subject [34–39], only a brief and informal description of the main concepts is given.

Using the pattern recognition approach, the subjects are divided into different classes, each one characterized by different features (dynamic, morphological, textural, clinical, spatiotemporal, pharmacokinetic). The classes were formed in such a way that individuals belonging to a given class were characterized by similar values, occupying a region in the multidimensional feature space, well separated from the other classes.

Pattern recognition methods are based on three main phases: feature extraction/selection, training, and classification. While the training and classification tasks can be considered a well-defined area, the extraction/selection of the most appropriate features for a specific field of research must be delineated specifically.

In the first phase, feature extraction means that existing features are combined to produce new ones. Several methods can be used to achieve this task. The main problem is that during the transformation (linear or non-linear feature combination), the physiological meaning of the original features may be lost.

In feature selection, only some of the features are chosen to eliminate redundant features, thus improving system efficiency. Two types of approach can be used for feature selection, namely wrapped and filter methods. A wrapper method uses a specific classifier to evaluate the features. This leads to high performance, since the selected features are the most appropriate for the chosen classifier. The filter method does not take into account the classifier. It is thus less computationally expensive when the number of features is very large.

After feature extraction/selection, testing is conducted, in which the classifier is designed using a training data set with the characteristics of the population under investigation. The classifier first has to be proven using a testing data set, which is different from the training data set. The performance of the classifier and its sensitivity to the training and testing data sets can be analyzed using two methods: the leave-one-out method and the 10-fold method.

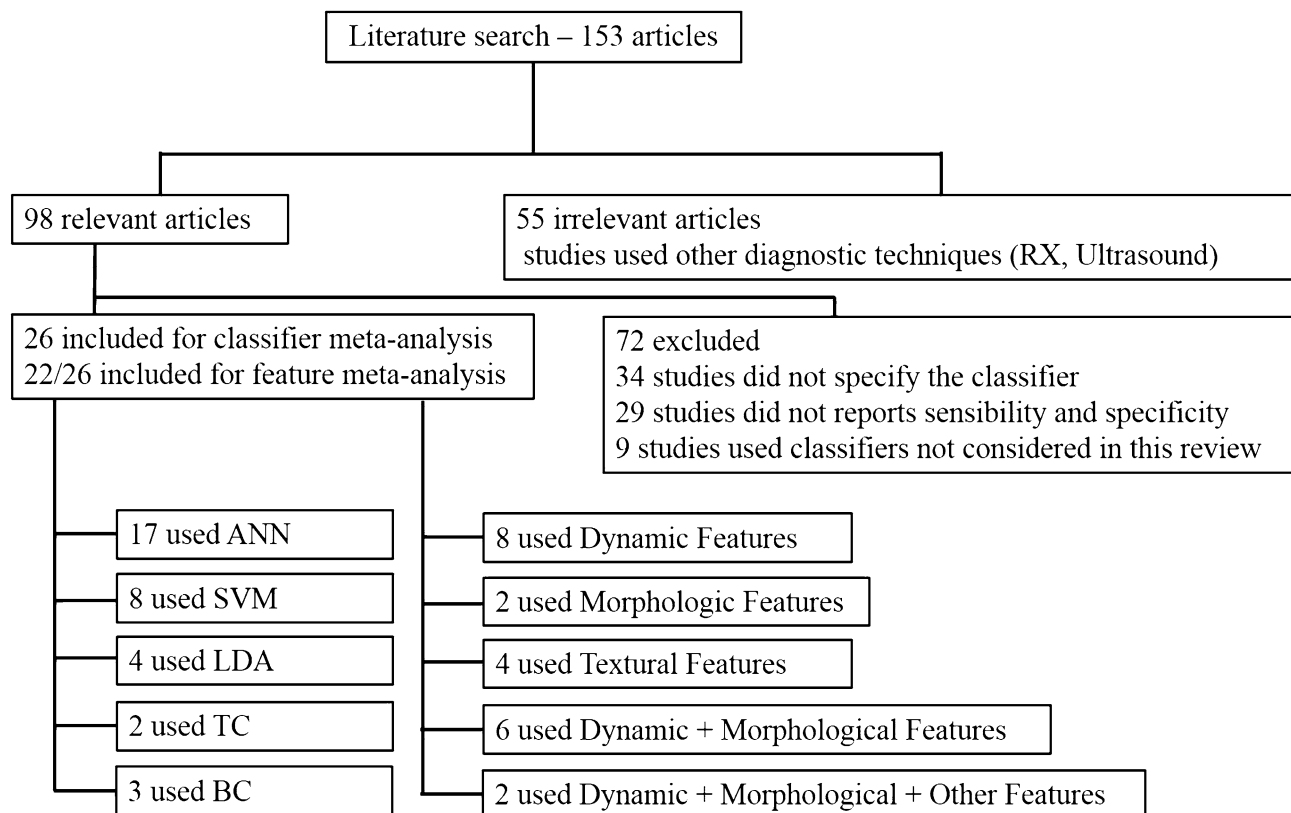
The leave-one-out method removes elements from the data set, one at a time. Then, the classifier is designed based on the remaining elements and is tested using the removed ones. This method can only be used on a small database. For the 10-fold method, the data set is divided into 10 subsets, and then a procedure similar to that of the leave-one-out method is applied.

For the final validation, a validation data set, which is different from both the training and testing data sets, is typically applied. The final phase is classification. Various machine learning algorithms can be used, such as ANN, SVM, LDA, TC, and BC.

A widely used strategy is to consider different classifiers at the same time. Each classifier receives the same set or subset of features as input and the final decision on the class is taken using an adequate scheme.

### 2.3 Classifiers

In this survey, we focus on the most commonly used classifiers, namely ANN, SVM, LDA, TC, and BC. A brief informal description of each classifier is given. The theoretical details of these classifiers can be found elsewhere [26–29].



**Fig. 1** Included and excluded studies in systematic review

### 2.3.1 Artificial Neural Networks

ANN is a set of mathematical models that mimic the behavior of neurons in the human brain, connected to each other through synapses. A neural network is a collection of elements (neurons) that are individually able to perform a fairly simple task and are interconnected with each other through unidirectional channels in order to perform more complex behavior. The output signal of the network is calculated on the basis of an input signal (feature vector) and the local memory of each neuron. An intermediate hidden layer of neurons is applied in classification problems that cannot be solved by a single-layer network. An input-hidden-output structure is called a multilayer perceptron (a single-layer network is called a perceptron). The set of inputs and the contents of the local memory are considered the inputs of a suitable transfer function that calculates the output, which will be propagated to other neurons, and so on, until it reaches the final output of the network [34–39]. This architecture is capable of drawing a hyperplane in the feature space that separates the classes. Typically, it achieves this task using an algorithm for updating weights called back propagation [34–39].

### 2.3.2 Support Vector Machines

SVM is a binary classifier that separates data using a hyperplane, determined based on selected points from the training set. While the traditional methods for classification are based on the minimization of empirical risk, or the optimization of performance on the training set, SVM minimizes the structural risk, i.e., the probability that new samples are classified correctly for a fixed probability distribution of the data.

Given a set of linearly separable data, there are various hyperplane separators that discriminate the data correctly. SVM identifies the hyperplane that, in addition to being correct for the training set, is also able to maximize the margin, defined as the sum of the distances between the hyperplane and the nearest points on both sides of it. Such points are called support vectors and are the only points of the training set used for determining the optimal hyperplane [34–36].

### 2.3.3 Linear Discriminant Analysis

LDA is a method of classification whose basic idea is to build decisional contours to separate the objects of the

**Table 1** Numbers of studies and patients per classifier

Classifier	Number of studies	Total number of patients
ANN	17	1960
SVM	8	949
LDA	4	133
TC	2	176
BC	3	343

classes using the optimization of the error criterion. The method is based on the fact that the distributions of data with a large variance between two classes and those with a smaller variance within each class are easy to separate [34–37].

### 2.3.4 Tree-Based Classifiers

A TC is based on the idea of dividing a complex decision into a union of many easier ones, so that at the end, the solution obtained reflects the desired one. This kind of classifier has the advantage of being very fast. A TC simplifies complex calculations and deletes unnecessary ones, and is also very intuitive and easily understandable [34–36].

### 2.3.5 Bayesian Classifiers

A Bayesian network is a probabilistic graphical model of knowledge in an uncertain domain that can be used to build a BC, which estimates statistical data from the training set to calculate the posterior probability. The main advantages of BC are a simple association between the knowledge of the model and that of the data, and ability to model reality in conditions of uncertainty [34–36].

## 2.4 Features for Breast Classification

### 2.4.1 Dynamic Features

Dynamic features (DYN) describe the temporal dynamics of the signal through measures obtained directly from the time-intensity curve. They are therefore model-free, since they are not calculated according to a model. The main dynamic features are area, maximum intensity ratio, relative enhancement, relative enhancement slope, basal signal, perfusion index, sum of intensities difference (SOD), wash-in, wash-out, and time to peak [41, 42, 49].

### 2.4.2 Pharmacokinetic Features

Pharmacokinetic features (PK) reflect some physiological parameters of tissues and are calculated on the basis of

mathematical models according to a model-based strategy [45–48]. They include extracellular extravascular space (EES), plasma space, and transfer constants between the plasma space and the EES. Moreover, when more complex kinetic models were used, pharmacokinetic features also include permeability flux, extraction fraction, and capillary transit time [45–48].

### 2.4.3 Spatiotemporal Features

Spatiotemporal features (STEP) model the signals in a four-dimensional space to capture not only the temporal dynamics and the architectural characteristics, but also the spatial variations of the voxels. Spatial and temporal properties are combined to obtain these features [17, 33].

### 2.4.4 Morphological Features

Morphological features (MOR) describe the shape and structure of the region of interest obtained in detection. The main morphological features are area, circularity, compactness, complexity, perimeter, radial length, smoothness, roughness, sphericity, eccentricity, volume, rectangularity, solidity, speculation, convexity, curvature, and edge [25, 32, 40, 41].

### 2.4.5 Textural Features

Textural features (TEX) are based on the texture of the image, i.e., its geometric structure. There are many definitions of texture; in general terms, it can be seen as a function of local spatial variation in the intensity of the voxels. Therefore, textural features replace the original values of the voxels with measures that describe their statistical properties: mean, median, standard deviation, kurtosis, and skewness [43, 44].

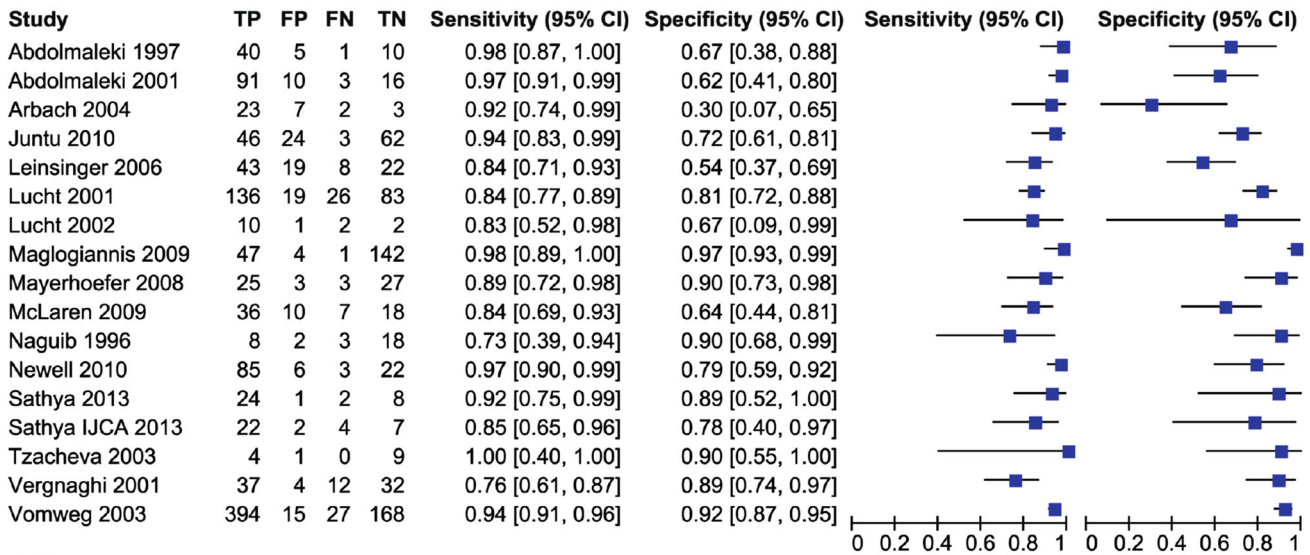
### 2.4.6 Clinical Features

Clinical features (CLI) relate to the patient's medical records and can provide additional information or instructions that may be useful for classification [31].

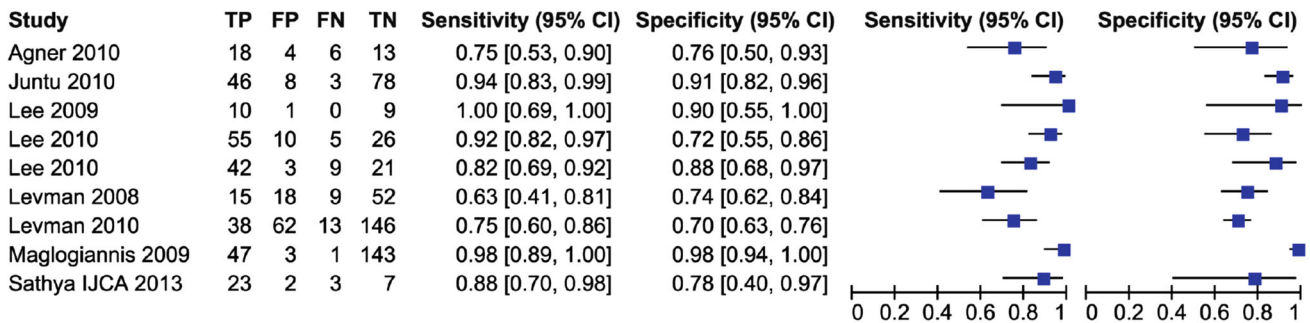
## 2.5 Data Analysis

All data analysis were performed using the software Rev-Man (version 5.2) [50]. Forest plots were constructed to graphically present the sensitivity and specificity values, with corresponding 95 % confidence intervals (CIs), for the individual studies. A summary receiver operating characteristic (sROC) curve was constructed using the same software.

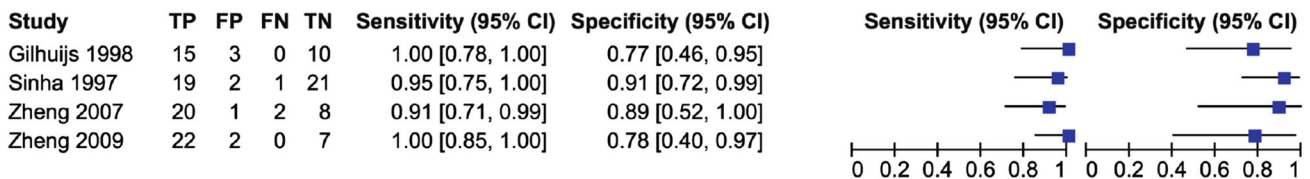
**ANN**



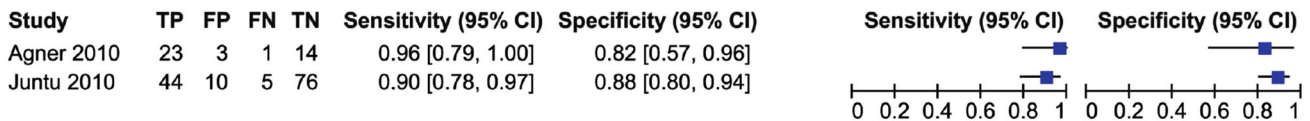
**SVM**



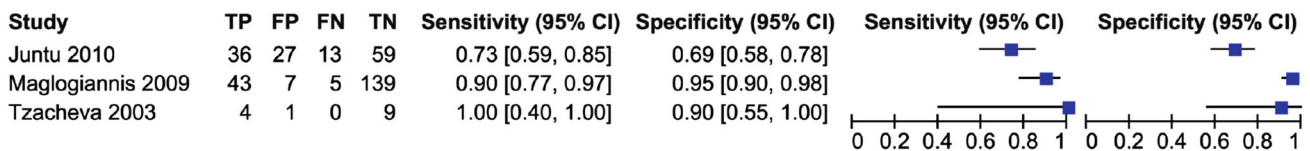
**LDA**



**TC**

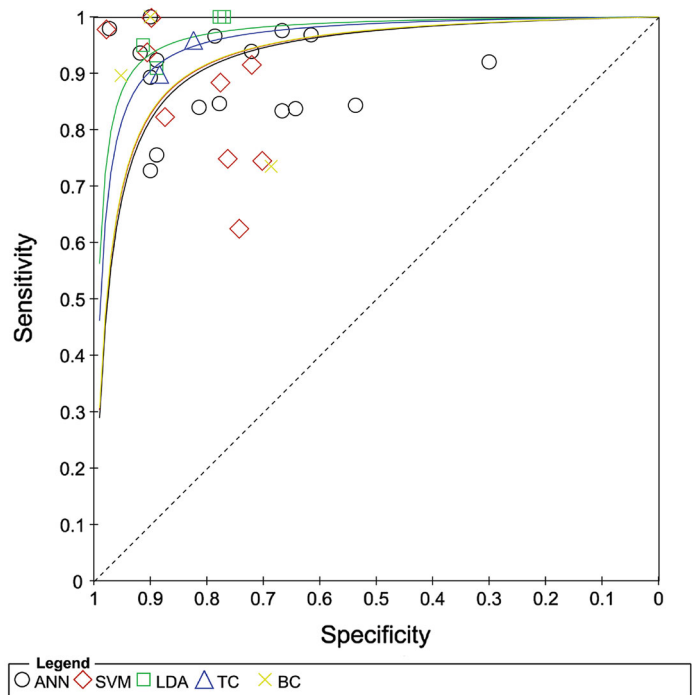


**BC**



**Fig. 2** Forest plot of sensitivity and specificity, with corresponding 95 % CIs, of included studies, divided by classifiers

**Fig. 3** Sensitivity and specificity plotted in receiver operating characteristic space for individual studies; sROC curves are plotted from data points for each classifier



**Table 2** Numbers of studies and patients per feature

Feature	Number of studies	Total number of patients
DYN	8	1000
MOR	2	49
TEX	4	668
DYN + MOR	6	930
DYN + MOR + other	2	125

### 3 Results

By using the search terms described earlier, we identified 153 studies from 1995 through 2014. Of these, 55 studies used diagnostic techniques other than DCE-MRI, 34 studies did not specify which classifiers were used, 29 did not have sufficient data (did not report sensitivity and specificity), and 9 studies used classifiers that were excluded from this review. 26 studies remained for inclusion in our meta-analysis, 20 were performed after 2000 (Fig. 1).

As shown in Table 1, the studies included in this review are divided by classifier as follows: 17 used ANN, 8 used SVM, 4 used LDA, 2 used TC, and 3 used BC. Some studies used multiple classifiers. The results of this first meta-analysis are shown in Figs. 2 and 3. Figure 2 shows the values of TP, FP, FN, TN, sensitivity, and specificity for each study, divided according to the applied classifier. Figure 3 shows the sROC curves for each classifier.

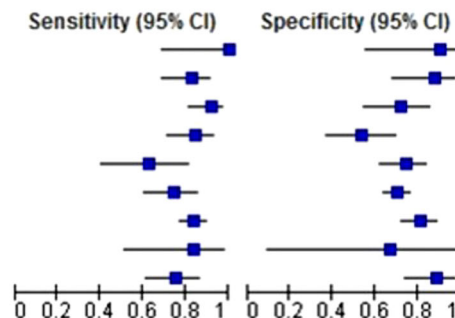
The included studies also considered different features. In this second systematic meta-analysis, we considered only the studies that reported a detailed description of used features (22 of the 26 included studies). The most used features were dynamic features, followed closely by morphological features, and then textural features (Table 2). The results of this second meta-analysis are shown in Figs. 4 and 5, which respectively show the forest plot and the sROC curves.

### 4 Discussion

In the last two decades, many studies have tackled the problem of tumor lesion classification based on DCE-MRI data analysis. It has been recognized that this problem can be addressed in a pattern recognition framework with the use of suitable features and classifiers. The present study performed a systematic review of several machine

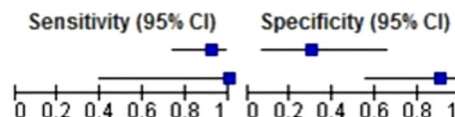
**Dyn**

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Lee 2009	10	1	0	9	1.00 [0.69, 1.00]	0.90 [0.55, 1.00]
Lee 2010	42	3	9	21	0.82 [0.69, 0.92]	0.88 [0.68, 0.97]
Lee 2010	55	10	5	26	0.92 [0.82, 0.97]	0.72 [0.55, 0.86]
Leinsinger 2006	43	19	8	22	0.84 [0.71, 0.93]	0.54 [0.37, 0.69]
Levman 2008	15	18	9	52	0.63 [0.41, 0.81]	0.74 [0.62, 0.84]
Levman 2010	38	62	13	146	0.75 [0.60, 0.86]	0.70 [0.63, 0.76]
Lucht 2001	136	19	26	83	0.84 [0.77, 0.89]	0.81 [0.72, 0.88]
Lucht 2002	10	1	2	2	0.83 [0.52, 0.98]	0.67 [0.09, 0.99]
Vergnaghi 2001	37	4	12	32	0.76 [0.61, 0.87]	0.89 [0.74, 0.97]



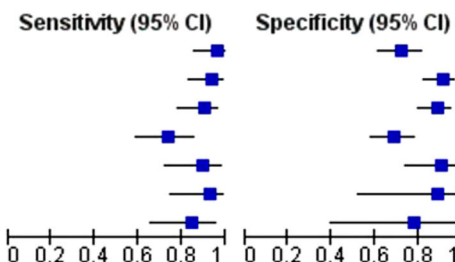
**Mor**

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Arbach 2004	23	7	2	3	0.92 [0.74, 0.99]	0.30 [0.07, 0.65]
Tzacheva 2003	4	1	0	9	1.00 [0.40, 1.00]	0.90 [0.55, 1.00]



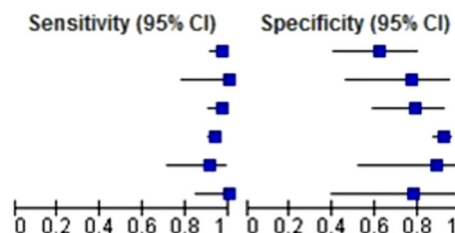
**Tex**

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Juntu 2010	46	24	2	62	0.96 [0.86, 0.99]	0.72 [0.61, 0.81]
Juntu 2010	46	8	3	78	0.94 [0.83, 0.99]	0.91 [0.82, 0.96]
Juntu 2010	44	10	5	76	0.90 [0.78, 0.97]	0.88 [0.80, 0.94]
Juntu 2010	36	27	13	59	0.73 [0.59, 0.85]	0.69 [0.58, 0.78]
Mayerhoefer 2008	25	3	3	27	0.89 [0.72, 0.98]	0.90 [0.73, 0.98]
Sathya 2013	24	1	2	8	0.92 [0.75, 0.99]	0.89 [0.52, 1.00]
Sathya IJCA 2013	22	2	4	7	0.85 [0.65, 0.96]	0.78 [0.40, 0.97]



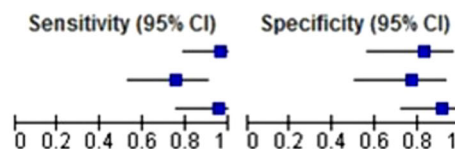
**Dyn+Mor**

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abdolmaleki 2001	91	10	3	16	0.97 [0.91, 0.99]	0.62 [0.41, 0.80]
Gilhuijs 1998	15	3	0	10	1.00 [0.78, 1.00]	0.77 [0.46, 0.95]
Newell 2010	85	6	3	22	0.97 [0.90, 0.99]	0.79 [0.59, 0.92]
Vornweg 2003	394	15	27	168	0.94 [0.91, 0.96]	0.92 [0.87, 0.95]
Zheng 2007	20	1	2	8	0.91 [0.71, 0.99]	0.89 [0.52, 1.00]
Zheng 2009	22	2	0	7	1.00 [0.85, 1.00]	0.78 [0.40, 0.97]



**Dyn+Mor+other**

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Agner 2010	23	3	1	14	0.96 [0.79, 1.00]	0.82 [0.57, 0.96]
Agner 2010	18	4	6	13	0.75 [0.53, 0.90]	0.76 [0.50, 0.93]
Sinha 1997	19	2	1	21	0.95 [0.75, 1.00]	0.91 [0.72, 0.99]

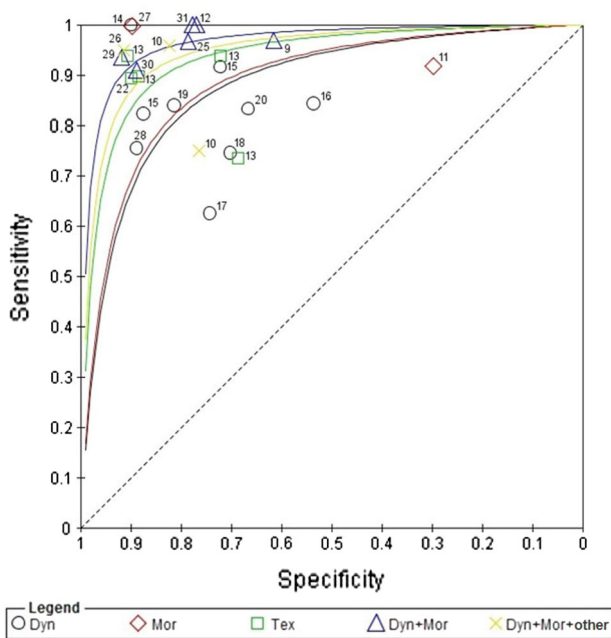


**Fig. 4** Forest plot of sensitivity and specificity, with corresponding 95 % CIs, of included studies, divided by features

learning algorithms proposed in the literature to classify breast lesion using several features categories such as dynamic, morphological, and textural features in DCE-MRI. Our results indicate that the choice of features does not affect the selection of the classifier; in fact, many authors used different combinations of features and classifiers, without compromising the validity of their study. We can thus safely say that the choice of features and that of classifier must not necessarily be related.

Although machine learning algorithms heavily depend on the training data and extracted features and the reported sensitivities and specificities cannot be compared directly, we can draw some conclusions. Based on the collected data, the average values of sensitivity and specificity were calculated for each classifier. The sensitivity and specificity were respectively 91 and 83 % for ANN, 85 and 82 % for SVM, 96 and 85 % for LDA, 92 and 87 % for TC, and 82 and 85 % for BC.





**Fig. 5** Sensitivity and specificity plotted in receiver operating characteristic space for individual studies; sROC curves are plotted from data points for each feature

The average values are 90 % for sensitivity and 83 % for specificity. At first, it might seem that LDA and TC have the best performance (Fig. 2); however, from Table 1, these two classifiers have the smallest samples of patients. In fact, LDA was used in only 4 studies, with a total of 133 patients, and TC was used in only 2 studies, with for a total of 176 patients.

In Fig. 3, the sROC curves corresponding to the SVM, ANN, and BC classifiers are in red, black, and yellow, respectively. These three curves are almost superimposed. A few studies achieved very good performance in term of both sensitivity and specificity. Considering all the studies involving ANN, SVM, and BC, these classifiers appear to have similar performances. However, these performance seem less precise than those seen previously; this probably stems from the fact that the statistical samples used in these cases are larger, and thus more accurate estimations of performance were obtained.

We now discuss the results of our second meta-analysis. From Fig. 4, we calculated the average values of sensitivity and specificity for each type of feature. The sensitivity and specificity are respectively 82 and 74 % for dynamic features, 93 and 60 % for morphological features, 88 and 81 % for textural features, 95 and 86 % for a combination of dynamic and morphological features, and 88 and 84 % for a combination of dynamic, morphological, and other features.

In Fig. 5, the sROC curves show that dynamic and morphological features used alone have similar

performances, but they are not precise. Better performance can be achieved using multiple features simultaneously. In Fig. 5, the least and most accurate curves are for studies that used only dynamic features and those that used a combination of dynamic and morphological features, respectively. Table 2 shows that 8 studies used only dynamic features, with a total of 1000 patients, and that 6 studies used a combination of dynamic and morphological features, with a total of 930 patients. The numbers of statistical samples in these cases are very similar and thus the performance improvement is associated with the use of more kinds of feature.

The contribution of literature surveys in general and of systematic reviews in particular, is to collect much information (in this case, types of classifiers, types of features, results on small and large populations, etc.) in a single place. This study adopted a systematic approach to summarize the performances (weighted on the basis of population size) of many studies, giving an overall indication of which methods can give good results. It must be underlined that even if a technique shows promising results for a small patient population, it must be verified with larger samples before it can become a standard protocol that clinicians can use in routine examinations.

## 5 Conclusion

This study performed a systematic review of several pattern analysis approaches for classifying breast lesions using dynamic, morphological, and textural features in (DCE-MRI) images. Our results indicate that LDA and TC have the best performance and that the remaining classifiers analyzed in this review (ANN, SVM, and BC) have similar performances, but are less precise than LDA and TC. This probably stems from the fact that the numbers of statistical samples used for the latter classifiers are larger, allowing a more accurate analysis. Moreover, dynamic and morphological features achieve better performance when used simultaneously in a given classifier.

One of the main issues that emerged from this study is the lack of standardization of the breast MRI exam. The scanning protocols for breast MRI vary in terms of pulse sequence parameters, spatial and temporal resolutions, field of view, exam duration, contrast agent dose injected, type of infusion, image pre- and post-processing, etc. This number of variables makes it difficult to compare studies. Efforts at an international level should be directed toward the assessment of guidelines.

A second issue related to the previous one is the lack of a publically available database of breast MRI images for the assessment of pattern recognition algorithms for feature extraction and classification. Such a database will improve

the evaluation of algorithms for image processing, feature extraction, and classification. Moreover, small sample sizes, which is an issue in a large number of the examined studies, could be addressed in this way.

A third issue is the variety of feature types used in previous studies. The majority of formulas used in previous studies try to extract information from the time course of the contrast agent and from the heterogeneity of the tumor. However, often, because of differences in MRI scanning protocols, the mathematical formulas of certain features cannot be directly used in different MRI settings. Moreover, blindly mixing features is not always a good approach because the mixtures cannot be properly understood by the clinician. An optimal set of features should be capable of effectively classifying tumors, be used in every MRI setting, and be as simple as possible to make sense for radiologists.

If the space of features is good, the project of the classifier would be simplified (e.g., linear or minimum distance). However, it is in general possible to obtain better performance by using more sophisticated classifiers (ANN, SVM, etc.) in combination with mixed features. The risk in this case is that an incomprehensible (to humans) set of features could emerge in combination with a strongly non-linear classifier. The adequacy of such a situation is questionable.

**Author Contributions** Guarantors of the integrity of the entire study: AP, RF, MS, CS; study concepts/study design or data acquisition or data analysis/interpretation: all authors; manuscript drafting or manuscript revision for important intellectual content: all authors; approval of final version of submitted manuscript: all authors; literature research: RF, MS, SF, GC, DMA; statistical analysis: RF, MS, CS; and manuscript editing: AP, RF, MS, CS.

#### Compliance with Ethical Standards

**Conflict of Interests** None of the authors have any competing interests.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- American Cancer Society. (2012) Breast cancer facts and figures 2011–2012. Atlanta: American Cancer Society, Inc. Retrieved May 25, 2013, from [www.cancer.org/acs/groups/content/@epidemiologysurveillance/documents/document/acspc-030975.pdf](http://www.cancer.org/acs/groups/content/@epidemiologysurveillance/documents/document/acspc-030975.pdf)
- Saslow, D., Boetes, C., Burke, W., Harms, S., Leach, M. O., Lehman, C. D., et al. (2007). American Cancer Society guidelines for breast screening with MRI as an adjunct to mammography. *CA Cancer Journal for Clinicians*, *57*(2), 75–89.
- Heywang-Köbrunner, S. H., Viehweg, P., Heinig, A., & Küchler, C. (1997). Contrast-enhanced MRI of the breast: accuracy, value, controversies, solutions. *European Journal of Radiology*, *24*(2), 94–108.
- Corsetti, V., Houssami, N., Ghirardi, M., et al. (2011). Evidence of the effect of adjunct ultrasound screening in women with mammography-negative dense breasts: interval breast cancers at 1 year follow-up. *European Journal of Cancer*, *47*, 1021–1026.
- Fusco, R., Sansone, M., Filice, S. & Petrillo, A. (2016) Breast contrast-enhanced MR imaging: semiautomatic detection of vascular map. *Breast Cancer*, *23*(2), 266–277.
- Kuhl, C. K., Mielcareck, P., Klaschik, S., Leutner, C., Wardelmann, E., Gieske, J., et al. (1999). Dynamic breast MR imaging: Are signal intensity time course data useful for differential diagnosis of enhancing lesions. *Radiology*, *211*(1), 101–110.
- Siegmann, K. C., Krämer, B., & Claussen, C. (2011). Current status and new developments in breast MRI. *Breast Care (Basel)*, *6*(2), 87–92.
- Nie, K., Chen, J.-H., Yu, H. J., Chu, Y., Nalcioğlu, O., & Su, M.-Y. (2008). Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI. *Academic Radiology*, *15*(12), 1513–1525.
- Ikeda, D. M., Hylton, N., Kinkel, K., Hochman, M. G., Kuhl, C., Kaiser, W. A., et al. (2001). Development, standardization, and testing of a lexicon for reporting contrast-enhanced breast magnetic resonance imaging. *Journal of Magnetic Resonance Imaging*, *13*(6), 889–895.
- Abdolmaleki, P., Buadu, L. D., Murayama, S., Murakami, J., Hashiguchi, N., Yabuuchi, H., et al. (1997). Neural network analysis of breast cancer from MRI findings. *Radiation Medicine*, *15*, 283–293.
- Abdolmaleki, P., Buadu, L. D., & Naderimansh, H. (2001). Feature extraction and classification of breast cancer on dynamic. *Cancer Letters*, *171*, 183–191.
- Agner, S. C., Soman, S., Libfeld, E., McDonald, M., Thomas, K., Englander, S., et al. (2010). Textural kinetics: A novel dynamic contrast-enhanced (DCE)-MRI feature for breast lesion classification. *Journal of Digital Imaging*, *24*, 446–463.
- Arbach, L., Stolpen, A., & Reinhardt, J. M. (2004). Classification of breast MRI lesions using a backpropagation neural network (BNN). *Biomedical Imaging*, *1*, 253–256.
- Gilhuijs, K. G. A., Giger, M. L., & Bick, U. (1998). Computerized analysis of breast lesions in three dimensions using dynamic magnetic-resonance imaging. *Medical Physics*, *25*, 1647–1654.
- Juntu, J., Sijbers, J., De Backer, S., Rajan, J., & Van Dyck, D. (2010). Machine learning study of several classifiers trained with texture analysis features to differentiate benign from malignant soft-tissue tumors in T1-MRI images. *Journal of Magnetic Resonance Imaging*, *31*, 680–689.
- Lee, S.H., Kim, J.H., Park, J.S., Jung, Y.S. & Moon, W.K. (2009) Characterizing time-intensity curves for spectral morphometric analysis of intratumoral enhancement patterns in breast DCE-MRI: Comparison between differentiation performance of temporal model parameters based on DFT and SVD. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (pp. 65–68). ISBI '09.
- Lee, S. H., Kim, J. H., Cho, N., Park, J. S., Yang, Z., Jung, Y. S., et al. (2010). Multilevel analysis of spatiotemporal association features for differentiation of tumor enhancement patterns in breast DCE-MRI. *Medical Physics*, *37*, 3940–3956.

18. Leinsinger, G., Schlossbauer, T., Scherr, M., Lange, O., Reiser, M., & Wismuller, A. (2006). Cluster analysis of signal-intensity time course in dynamic breast MRI: Does unsupervised vector quantization help to evaluate small mammographic lesions? *European Radiology*, *16*, 1138–1146.
19. Levman, J., Leung, T., Causer, P., Plewes, D., & Martel, A. L. (2008). Classification of dynamic contrast-enhanced magnetic resonance breast lesions by support vector machines. *IEEE Transactions on Medical Imaging*, *27*, 688–696.
20. Levman, J. & Martel, A.L. Computer-aided diagnosis of breast cancer from magnetic resonance imaging examinations by custom radial basis function vector machine. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 5577–5580).
21. Lucht, R. E. A., Knopp, M. V., & Brix, G. (2001). Classification of signal-time curves from dynamic MR mammography by neural networks. *Magnetic Resonance Imaging*, *19*, 51–57.
22. Lucht, R., Delorme, S., & Brix, G. (2002). Neural network-based segmentation of dynamic MR mammographic images. *Magnetic Resonance Imaging*, *20*, 147–154.
23. Maglogiannis, I., Zafiroopoulos, E., & Anagnostopoulos, I. (2009). An intelligent system for automated breast cancer diagnosis and prognosis using SVM based classifiers. *Applied Intelligence*, *30*(1), 24–36.
24. Mayerhoefer, M. E., Breitenhofer, M., Amann, G., & Dominkus, M. (2008). Are signal intensity and homogeneity useful parameters for distinguishing between benign and malignant soft tissue masses on MR images? Objective evaluation by means of texture analysis. *Magnetic Resonance Imaging*, *26*, 1316–1322.
25. McLaren, C. E., Nie, K., Chen, W.-P., Chen, J.-H., Nalcioğlu, O., & Su, M.-Y. L. (2009). Prediction of malignant breast lesions from MRI features: a comparison of Artificial Neural Network and logistic regression techniques. *Academic Radiology*, *16*, 842–851.
26. Naguib, R. N. G., Adams, A. E., Horne, C. H. W., Angus, B., Sherbet, G. V., & Lennard, T. W. J. (1996). The detection of nodal metastasis in breast cancer using neural network techniques. *Physiological Measurement*, *17*, 297–303.
27. Newell, D., Nie, K., Chen, J.-H., Hsu, C.-C., Yu, H. J., Nalcioğlu, O., et al. (2010). Selection of diagnostic features on breast MRI to differentiate between malignant and benign lesions using computer-aided diagnosis: differences in lesions presenting as mass and non-mass-like enhancement. *European Radiology*, *20*, 771–781.
28. Sinha, S., Lucas-Quesada, F. A., DeBruhl, N. D., Sayre, J., Farria, D., Gorczyca, D. P., et al. (1997). Multifeature analysis of Gd-enhanced MR images of breast lesions. *Journal of Magnetic Resonance Imaging*, *7*(6), 1016–1026.
29. Tzacheva, A. A., Najarian, K., & Brockway, J. P. (2003). Breast cancer detection in gadolinium-enhanced MR images by static region descriptors and neural networks. *Journal of Magnetic Resonance Imaging*, *17*, 337–342.
30. Vergnaghi, D., Monti, A., Setti, E., & Musumeci, R. (2001). A use of a neural network to evaluate contrast enhancement curves in breast magnetic resonance images. *Journal of Digital Imaging*, *14*, 58–59.
31. Vomweg, T. M., Buscema, M., Kauczor, H. U., Teifke, A., Intraligi, M., Terzi, S., et al. (2003). Improved artificial neural networks in prediction of malignancy of lesions in contrast-enhanced MR-mammography. *Medical Physics*, *9*, 2350–2359.
32. Zheng, Y., Baloch, S., Englander, S., Schnall, MD. & Shen, D. (2007) Segmentation and classification of breast tumor using dynamic contrast-enhanced MR images. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, vol. 10(Pt 2) (pp. 393–401). Springer, Berlin
33. Zheng, Y., Englander, S., Schnall, M. D., & Shen, D. (2009). STEP: spatial-temporal enhancement pattern, for MR-based breast tumor diagnosis. *Medical Physics*, *36*, 3192–3204.
34. Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*. Elsevier: Academic Press.
35. Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification*. Hoboken: Wiley Interscience, Wiley.
36. Theodoridis, S. (2003). *Konstantinos Koutroumbas Pattern recognition*. Elsevier: Academic Press.
37. Fusco, R., Sansone, M., Filice, S., Granata, V., Catalano, O., Amato, D. M., et al. (2015). Integration of DCE-MRI and DW-MRI quantitative parameters for breast lesion classification. *BioMed Research International*, *2015*, 237863.
38. Sathya, J., & Geetha, K. (2013). Mass classification in breast DCE-MR images using an artificial neural network trained via a bee colony optimization algorithm. *ScienceAsia*, *39*, 294–306.
39. Sathya, J., & Geetha, K. (2013). Experimental investigation of classification algorithms for predicting lesion type on breast DCE-MR images. *International Journal of Computer Applications*, *8*(4), 1.
40. Ikeda, D. M., Hylton, N. M., Kinkel, K., Hochman, M. G., Kuhl, C. K., Kaiser, W. A., et al. (2001). Development, standardization, and testing of a lexicon for reporting contrast-enhanced breast magnetic resonance imaging studies. *Journal of Magnetic Resonance Imaging*, *13*, 889–895.
41. Fusco, R., Sansone, M., Petrillo, A., & Sansone, C. (2012). A multiple classifier system for classification of breast lesions using dynamic and morphological features in DCE-MRI. *Lecture Notes in Computer Science*, *7626*, 684–692.
42. Degenhard, A., Tanner, C., Hayes, C., Hawkes, D. J., & Leach, M. O. (2002). Comparison between radiological and artificial neural network diagnosis in clinical screening. *Physiological Measurement*, *23*, 727–739.
43. Castellano, G., Bonilha, L., Li, L. M., & Cendes, F. (2004). Texture analysis of medical images. *Clinical Radiology*, *59*, 1061–1069.
44. Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, *6*, 610–621.
45. Brix, G., Kiessling, F., Lucht, R., Darai, S., Wasser, K., Delorme, S., et al. (2004). Microcirculation and microvasculature in breast tumors: Pharmacokinetic analysis of dynamic MR image series. *Magnetic Resonance in Medicine*, *52*, 420–429.
46. Sansone, M., Fusco, R., Petrillo, A., Petrillo, M., & Bracale, M. (2011). An expectation-maximisation approach for simultaneous pixel classification and tracer kinetic modelling in dynamic contrast enhanced-magnetic resonance imaging. *Medical & Biological Engineering & Computing*, *49*(4), 485–495.
47. Fusco, R., Sansone, M., Maffei, S., Raiano, N., & Petrillo, A. (2012). Dynamic contrast-enhanced MRI in breast cancer: A comparison between distributed and compartmental tracer kinetic models. *Journal of Biomedical Graphics and Computing*, *2*(2), 23–36.
48. Fusco, R., Sansone, M., Petrillo, M., & Petrillo, A. (2014). Influence of parameterization on tracer kinetic modeling in DCE-MRI. *Journal of Medical and Biological Engineering*, *34*(2), 157–163.
49. Fusco, R., Sansone, M., Sansone, C. & Petrillo, A. (2012) Segmentation and classification of breast lesions using dynamic and textural features in dynamic contrast enhanced-magnetic resonance imaging. In *25th International Symposium on Computer-Based Medical Systems* (pp. 1–4).
50. Available online at link <http://ims.cochrane.org/revman/download>