Beyond breast density: a review on the advancing role of parenchymal texture analysis in breast cancer risk assessment

Aimilia Gastounioti, Emily F. Conant and Despina Kontos *

Abstract

Background: The assessment of a woman's risk for developing breast cancer has become increasingly important for establishing personalized screening recommendations and forming preventive strategies. Studies have consistently shown a strong relationship between breast cancer risk and mammographic parenchymal patterns, typically assessed by percent mammographic density. This paper will review the advancing role of mammographic texture analysis as a potential novel approach to characterize the breast parenchymal tissue to augment conventional density assessment in breast cancer risk estimation.

Main text: The analysis of mammographic texture provides refined, localized descriptors of parenchymal tissue complexity. Currently, there is growing evidence in support of textural features having the potential to augment the typically dichotomized descriptors (dense or not dense) of area or volumetric measures of breast density in breast cancer risk assessment. Therefore, a substantial research effort has been devoted to automate mammographic texture analysis, with the aim of ultimately incorporating such quantitative measures into breast cancer risk assessment models. In this paper, we review current and emerging approaches in this field, summarizing key methodological details and related studies using novel computerized approaches. We also discuss research challenges for advancing the role of parenchymal texture analysis in breast cancer risk stratification and accelerating its clinical translation.

Conclusions: The objective is to provide a comprehensive reference for researchers in the field of parenchymal pattern analysis in breast cancer risk assessment, while indicating key directions for future research.

Keywords: Breast cancer risk, Parenchymal texture analysis, Quantitative breast imaging, Digital mammography

Background

The incidence and mortality rates of breast cancer remain extremely high despite advances in screening and treatment [1]. In the USA, it is estimated that, in 2016, there will be 246,660 new cases of invasive breast cancer and 40,450 breast cancer deaths [2]. Therefore, better strategies are urgently needed to identify women at high risk for developing breast cancer who could benefit the most from supplemental screening and preventive therapies [3, 4].

Unfortunately, to date, the broadly available risk assessment models cannot identify high-risk women reliably within the general population. Current models predict either the risk of carrying a high-risk genetic mutation such as BRCA1/2 (e.g., Claus model, BOADICEA, and BRCAPRO) or the risk of developing breast cancer over time with or without such a mutation (e.g., Gail model, BOADICEA, Rosner-Colditz model, and Tyrer-Cuzick model) [5]. These models have only modest discriminatory capacity and continuing efforts are needed to improve these models at the individual level [6]. In addition, genetic susceptibility models are only useful in the familial setting (where cancer pedigree history is known) and are not of relevance to the general population where the great majority of women have no relevant family history. Therefore, in striving to tailor breast cancer screening recommendations for the individual woman [7] it is crucial to develop more accurate risk assessment models that can be easily adopted in routine clinical practice.
While mammography remains the cornerstone of early breast cancer detection [8], it also provides a readily accessible method to assess the distribution of fatty and dense, or fibroglandular (stromal and epithelial), tissues in the breast. In x-ray imaging, fatty tissue appears radiographically lucent, or darker, and dense tissue is radio-opaque, or brighter. Mammographic percent density (PD), a measure of the relative amount of fibroglandular tissue within the breast, has been shown to be related to screening sensitivity and specificity and has also been established as a strong independent risk factor for breast cancer [9–12]. Studies have repeatedly shown significant associations with breast cancer risk for both qualitative and quantitative breast density measures and a potential to improve cancer risk assessment models [13, 14]. Recent legislation in several US states mandates notification of breast density [15], and substantial research continues to be devoted to accurate measurement of this key biomarker and to its incorporation into risk prediction models [9, 16].

Compared to the global image measure of breast density, parenchymal texture descriptors can provide more refined, localized descriptors to characterize the complexity as well as the morphological distribution of the breast parenchymal patterns. Breast density measures are generally dichotomous, or each area or voxel of breast density measured in the mammogram is compared to a threshold of “dense” or “not dense” without reflecting the broader range and spatial distribution of the various breast parenchymal elements. Parenchymal textural features have been proposed as not only imaging markers that could identify parenchymal changes associated with breast cancer development [17–19], but also with subtypes and grading of subsequent breast malignancies [20–22]. In addition, there is growing evidence in support of textural features of the breast parenchyma reflecting inherent, independent, biologic risk factors associated with cancer development, and this may thus have the potential to augment breast density in assessing an individual woman’s risk of developing cancer [23–25]. Therefore, efforts to incorporate breast parenchymal texture analyses in breast cancer risk assessment have recently also gained substantial momentum.

This article reviews approaches to quantitate mammographic textural features and methods to incorporate these features into breast cancer risk assessment models, focusing primarily on novel computerized approaches. A systematic review of the literature in PubMed was performed to identify all original articles published up to April 2016 that evaluated computational measures of mammographic texture in breast cancer risk assessment. The following keywords were used in combination: “texture” or “parenchymal patterns” or “image features”, “mammography” or “mammogram”, and “breast cancer risk” or “mammographic risk”. To broaden the search, the “related articles” function provided in PubMed was also used, and all articles and citations obtained were reviewed. The references from all the articles identified were also examined for further relevant studies. The last search was conducted on 29 April 2016. Studies not considered relevant to the scope of the review were excluded; other exclusion criteria included: study not published in the English language, full text not available, letter to the editor, and duplicate publication. In the rest of this manuscript, we summarize key methodological details and evaluation results from the 44 research papers identified by the search and discuss future challenges in this promising research field.

**Mammographic texture analysis using automatically extracted features**

The value of characterizing the mammographic texture of the breast parenchyma in breast cancer risk estimation was originally demonstrated in the pioneering studies of Wolfe [26, 27], Boyd et al. [28–30], Gram et al. [31], and Brisson et al. [32], proposing visually assessed, qualitative or quantitative classifications which were based on the extent and the characteristics of breast densities in a mammogram. These early approaches have been used by several groups, generally reporting elevated risks among women with more complex parenchymal tissue patterns [33–45]. Nevertheless, these studies also observed increased heterogeneity and low reproducibility in corresponding risk estimates due to subjectivity and inter-observer variation in visual appraisal of the mammogram [33–45]. By introducing computerized texture features to automate the characterization of breast parenchymal patterns, later studies addressed the limitations of visual classifications and re-established the potential of texture descriptors in breast cancer risk assessment [46–50]. Since then, this research field has continuously been evolving. A variety of quantitative methodologies have been developed, involving different techniques to sample the breast and multiple texture descriptors to characterize the texture properties of the sampled regions of interest (ROIs) from cranio-caudal (CC) and mediolateral-oblique (MLO) view mammograms (Table 1).

In most studies, texture analysis has been performed within a single ROI in the breast (Table 1). This single ROI is usually placed in the retroareolar breast area, while, in some cases, it can be a larger region corresponding to the entire breast or to the largest rectangular box inscribed within the breast (Fig. 1a and b). In an attempt to capture the granularity and heterogeneity of the parenchymal texture within the breast, more recent studies have estimated texture in multiple ROIs throughout the breast (Fig. 1c and d). A lattice-based strategy
Table 1  Key studies in automated parenchymal texture analysis for breast cancer risk assessment

| Study                                                                 | Participating Institutions                                                                 | Mammograms | Dataset | Breast sampling | Texture features |
|-----------------------------------------------------------------------|-------------------------------------------------------------------------------------------|------------|---------|-----------------|-----------------|
|                                                                       |                                                                                           | F/D View   | A       | B               | S1   | S2 | T1   | T2   | T3   | T4   | T5   |
| Distinguishing or predicting cancer cases from controls                |                                                                                           |            |         |                 |      |     |      |      |      |      |      |
| Byng et al. (1997) [60]                                                | University of Toronto, Sunnybrook Health Science Centre, Ontario Cancer Institute          | F CC       | 354     | 354            | x    | x  | x    |      |      |      |      |
| Torres-Mejia et al. (2005) [72]                                        | LSHTM, Guy’s Hospital, UNAM, IPOFG                                                         | F CC/MLO   | 111     | 3100           | x    | x  |      |      |      |      |      |
| Wu et al. (2008) [76]                                                  | University of Michigan                                                                     | F CC       | 128     | 549            | x    |     |      |      |      |      | x    |
| Manduca et al. (2009) [66]                                             | Mayo Clinic, Moffitt                                                                      | F CC/MLO   | 246     | 522            | x    | x  | x    | x    | x    |      |      |
| Wei et al. (2011) [73]                                                 | University of Michigan                                                                     | F CC       | 136     | 246            | x    |     |      |      |      |      | x    |
| Nielsen et al. (2011) [61]                                            | University of Copenhagen, Nordic Bioscience, Delft University of Technology, RadboudUMC, Mayo Clinic | F MLO      | 245     | 250            | x    |     |      |      |      |      | x    |
| Brandt et al. (2011) [74]                                              | University of Copenhagen, RadboudUMC, Synarc Imaging Technologies                         | F MLO      | 245     | 245            | x    |     |      |      |      |      |      |
| Häberle et al. (2012) [56]                                            | Erlangen University Hospital, Fraunhofer Institute for Integrated Circuits IIS, IMPRS, UCLA | F CC       | 864     | 418            | x    | x  | x    | x    | x    |      |      |
| Li et al. (2012) [84]                                                  | University of Chicago                                                                     | D CC       | 75      | 328            | x    | x  | x    | x    |      |      |      |
| Chen et al. (2014) [75]                                               | University of Manchester                                                                  | D MLO      | 50      | 50             | x    |     |      |      |      |      |      |
| Nielsen et al. (2014) [64]                                            | University of Copenhagen, Nordic Bioscience, Biomedical, RadboudUMC, Mayo Clinic          | F CC/MLO   | 471     | 692            | x    |     |      |      |      |      |      |
| Li et al. (2014) [71]                                                 | University of Chicago                                                                     | D CC       | 75      | 328            | x    | x  | x    | x    |      |      |      |
| Karemore et al. (2014) [89]                                           | University of Copenhagen, RadboudUMC                                                      | F MLO      | 245     | 250            | x    |     |      |      |      |      |      |
| Zheng et al. (2015) [51]                                              | University of Pennsylvania                                                                | D MLO      | 106     | 318            | x    | x  | x    | x    |      |      |      |
| Sun et al. (2015) [53]                                                | University of Texas, China Northeastern University, TTUHS, Guanajuato University           | D CC       | 141     | 199            | x    | x  | x    |      |      |      |      |
| Tan et al. (2015) [77]                                                | University of Texas, University of Oklahoma, University of Pittsburgh                     | D CC/MLO   | 812     | 1084           | x    |     | x    | x    |      |      |      |
| Tan et al. (2015) [78]                                                | University of Oklahoma, University of Pittsburgh                                          | D CC/MLO   | 430     | 440            | x    |     | x    | x    | x    |      |      |
| Predicting the risk of carrying a high-risk genetic mutation           |                                                                                           |            |         |                 |      |     |      |      |      |      |      |
| Huo et al. (2000) [80]                                                | University of Chicago                                                                     | F CC       | 15      | 143            | x    | x  | x    |      |      |      |      |
| Huo et al. (2002) [55]                                                | University of Chicago, University of Pennsylvania                                         | F CC       | 30      | 142            | x    | x  | x    |      |      |      |      |
| Li et al. (2004) [54]                                                 | University of Chicago                                                                     | F CC       | 30      | 60             | x    | x  | x    | x    |      |      |      |
| Li et al. (2005) [81]                                                 | University of Chicago                                                                     | F CC       | 30      | 142            | x    | x  | x    | x    |      |      |      |
| Li et al. (2007) [82]                                                 | University of Chicago                                                                     | F CC       | 30      | 142            | x    |     |      |      |      |      | x    |
| Li et al. (2008) [83]                                                 | University of Chicago                                                                     | F CC       | 30      | 142            | x    |     |      |      |      |      |      |
| Li et al. (2012) [84]                                                 | University of Chicago                                                                     | D CC       | 53      | 328            | x    | x  | x    | x    |      |      |      |
| Li et al. (2014) [71]                                                 | University of Chicago                                                                     | D CC       | 53      | 328            | x    |     | x    | x    |      |      |      |
| Gierach et al. (2014) [85]                                            | University of Chicago, NCI-NIH, Washington Radiology Associates, Genentech, USUHS, UCL, WRNMC, Westat Inc. | F CC       | 137     | 100            | x    | x  | x    | x    |      |      |      |

The Table describes the image data used in each study, including type of mammograms and dataset size, as well as methodological details for the computerized texture analysis, the technique of breast sampling, and algorithm implementation of texture features. IMPRS International Max Planck Research School for Optics and Imaging, IPOFG Instituto Portugues de Oncologia Francisco Gentil, LSHTM London School of Hygiene and Tropical Medicine, Moffitt Moffitt Cancer Center and Research Institute, NCI-NIH National Cancer Institute, National Institutes of Health, RadboudUMC Radboud University Nijmegen Medical Centre, TTUHS Texas Tech University Health Sciences, UCL University College London, UCLA University of California at Los Angeles, UNAM Universidad Nacional Autonoma de Mexico, USUHS Uniformed Services University of the Health Sciences, WRNMC Walter Reed National Military Medical Center.

Mammograms: F Digitized screen-film, D Full-field digital, CC cranio-caudal, MLO mediolateral-oblique; Dataset: A cancer cases (prior, unaffected, images from the contralateral breast or the entire breast/dense tissue as a single region of interest (ROI), S2 multiple ROIs covering the entire breast; Types of texture features: T1 gray-level histogram, T2 co-occurrence, T3 run-length, T4 structural/pattern, T5 multi-resolution/spectral.
which splits the entire breast into multiple square patches was proposed by Zheng et al. [51, 52] showing that, with respect to single ROI methodologies, this breast sampling technique may improve risk assessment, with performance being maximized when smaller patches (6.3 × 6.3 mm²) are used. Multiple ROIs defined at various scales of breast tissue density were used by Sun et al. [53], where it was shown that fusing features from different density scales may prove to be another effective way to enhance the cancer prediction performance.

The texture descriptors used in breast cancer risk assessment to date can be broadly classified into five feature groups (Table 2), each of which reveals different aspects of the mammographic texture (Fig. 2): 1) grey-level intensity/histogram features [54–56]; 2) co-occurrence (Haralick/Markovian) descriptors [57]; 3) run-length features [58, 59]; 4) structural/pattern measures [46, 60–65]; and 5) multi-resolution/spectral features [53, 61, 64, 66, 67]. Gray-level intensity histogram features are common first-order statistics which describe the distribution of gray-level intensity within the breast tissue. The co-occurrence features also consider the spatial relationships of pixel intensities in different directions and are based on the gray-level co-occurrence matrix (GLCM) which encodes the relative

| Table 2 Parenchymal texture descriptors for breast cancer risk assessment; texture descriptors which have been examined in association with breast cancer risk, classified to five feature groups |
|---|
| Grey-level histogram features [51, 53–56, 60, 66, 71, 72, 77, 80, 81, 84, 85] |
| min intensity | skewness | 5th percentile | energy |
| max intensity | kurtosis | 5th percentile mean | root mean square variation |
| standard deviation | entropy | 95th percentile |
| mean intensity | sum intensity | 95th percentile mean |
| Co-occurrence features [51, 53–56, 66, 71, 77, 78, 80, 81, 84, 85] |
| cluster shade | entropy | inverse difference moment | difference entropy |
| correlation | inertia | sum variance | homogeneity |
| Haralick correlation | difference moment | sum average | product moment |
| energy | coarseness | difference variance | triangular symmetry |
| Run-length measures [51, 56, 66, 73, 76–78] |
| long run emphasis | gray-level non-uniformity | high gray level run emphasis | run percentage |
| short run emphasis | run-length non-uniformity | low gray level run emphasis | number of runs |
| Structural/PATTERN measures [51, 54, 56, 60, 61, 64, 66, 72, 74, 77, 78, 81, 82, 84, 85, 89] |
| fractal dimension | local binary pattern | Hessian matrix | Weber local descriptors |
| lacunarity | Law’s masks | edge enhancing index | directional gradient |
| Multi-resolution/Spectral features [53–56, 61, 64, 66, 71, 75, 78, 80, 81, 83–85, 89] |
| Fourier power spectrum | wavelet/Gabor | Gaussian Kernels | power-law spectrum |
frequency of neighboring intensity values. Run-length features capture the coarseness of texture in specified directions by measuring strings of consecutive pixels (i.e., runs) which have the same gray-level intensity along specific linear orientations. Fine textures tend to contain more short runs with similar gray-level intensities, while coarse textures have longer runs with different gray-level intensities. Structural features capture the architectural composition of the parenchyma by characterizing the tissue complexity, the directionality of flow-like structures in the breast, and intensity variations between central and neighboring pixels. Finally, multi-resolution/spectral features use spatial frequency transforms, such as Fourier, wavelet/Gabor, and the Power spectrum, to characterize intrinsic periodic texture structures that repeat over multiple scales.

**Towards a new breast cancer risk assessment paradigm based on mammographic texture descriptors**

The proposed methodologies have been applied primarily to digitized film-screen mammograms and more recently on full-field digital mammograms. Texture descriptors have been evaluated in a few prospective and a larger number of retrospective case–control studies, where their discriminatory capacity in breast cancer prediction was typically assessed in terms of the area under the ROC curve (AUC) measuring their ability to distinguish between cancer cases and controls (Table 3). The potential of mammographic texture in breast cancer risk assessment has also been investigated in studies with BRCA1/2 mutation carriers, where the AUC was evaluated in terms of the performance of the texture features in predicting a woman’s risk of carrying this high-risk genetic mutation. Although hereditary breast cancers account for 5–10% of incident breast cancers, women who inherit a mutated form of the BRCA1/2 gene have up to 87% risk of developing breast cancer by the age of 70 years [68]. As such, and considering that mammographic PD has not been associated with BRCA1/2 mutation status [69–71], the ability of texture to identify potential BRCA1/2 carriers could have important value in risk stratification.

**Associations of parenchymal texture with breast cancer in case–control studies**

Byng et al. [60] were the first to evaluate automatically calculated parenchymal texture descriptors directly as independent risk factors for breast cancer. The authors reported on data from a prospective case–control study using 354 incident cases diagnosed with histologically verified invasive breast carcinoma at least 1 year after their entry in the Canadian National Breast Screening Study, and 354 age-matched controls with at least 7 years of negative follow-up. Two grey-level intensity histogram texture features were estimated in screen-film mammograms; specifically, skewness averaged over individual 6.2 × 6.2 mm² patches in the breast and the fractal dimension estimated by considering the entire breast as a single ROI. For both features, the results showed moderate relative risk (RR) after adjustments for the effects of other risk factors, i.e., age at menarche, menopausal status, age at first time pregnancy, number of live births, family history of breast carcinoma, height, and weight (RR = 3.35 and RR = 3.35 for skewness and fractal dimension, respectively), while no
| Study | Participating Institutions | Dataset | Model | Discriminatory capacity (AUC) |
|-------|----------------------------|---------|-------|------------------------------|
|       |                            | A  B  m |       |                              |
| Distinguishing between cancer cases and healthy women |                                 |   |     |                              |
| Wu et al. (2008) [76] | University of Michigan | 128 549 No | LDA<sub>CV</sub> | 0.73 |
| Manduca et al. (2009) [66] | Mayo Clinic, Moffitt | 246 522 Yes | LR<sub>CV</sub> | [0.58, 0.60] 0.58 |
| Wei et al. (2011) [73] | University of Michigan | 136 246 No | LDA | 0.74* 0.61 0.76 |
| Nielsen et al. (2011) [61] | University of Copenhagen, Nordic Bioscience, Delft University of Technology, RadboudUMC, Mayo Clinic | 245 250 No | LR<sub>CV</sub> | 0.63 0.60 0.66* |
| Brandt et al. (2011) [74] | University of Copenhagen, RadboudUMC, Synarc Imaging Technologies | 245 245 Yes | kNN<sub>CV</sub> | 0.63 0.56 |
| Häberle et al. (2012) [56] | Erlangen University Hospital, Fraunhofer Institute for Integrated Circuits IIS, IMPRS, UCLA | 864 418 Yes | LR<sub>CV</sub> | 0.75 0.51 0.75 |
| Li et al. (2012) [84] | University of Chicago | 75 328 No | BANN<sub>CV</sub> | 0.73 |
| Chen et al. (2014) [75] | University of Manchester | 50 50 No | LR | 0.71 0.62 0.68 |
| Nielsen et al. (2014) [64] | UCPH, Nordic Bioscience, Biomediq, RadboudUMC, Mayo Clinic | 245 250 No | LR<sub>CV</sub> | 0.63 0.60 0.66* |
| Nielsen et al. (2014) [64] | UCPH, Nordic Bioscience, Biomediq, RadboudUMC, Mayo Clinic | 226 442 Yes | LR<sub>CV</sub> | 0.63 0.60 0.66* |
| Li et al. (2014) [71] | University of Chicago | 67 268 Yes | BANN<sub>CV</sub> | 0.70* 0.57 0.68 |
| Karemne et al. (2014) [80] | UCPH, RadboudUMC | 245 250 Yes | kNN<sub>CV</sub> | 0.59 |
| Zheng et al. (2013) [51] | University of Pennsylvania | 106 318 Yes | LR<sub>CV</sub> | 0.85* 0.59 0.86 |
| Sun et al. (2013) [53] | University of Texas, China Northeastern University, University of Oklahoma, UT-HS, Guting Medical University | 141 199 No | SVM<sub>CV</sub> | 0.73 |
| Tan et al. (2015) [77] | University of Texas, University of Oklahoma, University of Pittsburgh | 812 1084 No | ANN<sub>CV</sub> | 0.71 |
| Tan et al. (2015) [78] | University of Oklahoma, University of Pittsburgh | 430 440 No | ANN<sub>CV</sub> | [0.64, 0.73] |
| Predicting the risk of carrying a high-risk genetic mutation |                                 |   |     |                              |
| Huo et al. (2000) [80] | University of Chicago | 15 143 No | LDA | [0.59, 0.82] |
| Huo et al. (2000) [80] | University of Chicago | 15 30 Yes | LDA | [0.53, 0.87] |
| Huo et al. (2002) [55] | University of Chicago, University of Pennsylvania | 30 142 No | LDA | 0.91 |
| Huo et al. (2002) [55] | University of Chicago, University of Pennsylvania | 30 60 Yes | LDA | 0.92 |

Age, BMI, family history of breast cancer, # of previous breast biopsies

Age, BMI, family history of breast cancer, parity, age at first term pregnancy

Age, BMI, menopause, hormonal use

Age, BMI, menopause, hormonal use

Age, BMI, family history of breast cancer, hormonal use, age at first term pregnancy
Additional contribution to risk was found in models that incorporated breast density measures. Similar conclusions were reported by Torres-Mejia et al. [72] who estimated the same texture features and lacunarity, a measure of the degree of structural variation in image intensities within the breast, from prospectively collected data of 111 breast cancer cases and 3100 controls.

The promising results of these early studies were followed by retrospective studies using more complex parenchymal texture descriptors [61, 64, 73–76]. Wei et al. [73] investigated the associations of breast cancer risk with run-length features, using two different implementations of run-length statistics: namely, the conventional approach for calculating the runs of pixels in one direction and an extension for the two-dimensional space [76]. The authors found that the run-length measures calculated in the retroareolar region of the breast could serve as an additional risk factor that could not be explained by established breast cancer risk factors (i.e., age, BMI, family history of breast cancer, and number of previous biopsies) and breast density. A mammographic texture resemblance (MTR) marker based on multi-scale Gaussian features was proposed by Nielsen et al. [61]. This marker demonstrated high case-control discriminatory performance (AUC = 0.60–0.63) in two independent cohorts within the Dutch screening program [61] and the Mayo Mammography Health Study [61, 64], while performance was optimized by an aggregate marker combining MTR with density measures (AUC = 0.66). Gaussian derivative features at multiple scales were examined in a cross-sectional study with MLO-view film mammograms of 245 cancer cases and 245 controls from the Nijmegen risk-assessment study [74]. In this work, derivative features were extracted using an anatomically oriented breast coordinate system and, compared to breast PD, demonstrated enhanced breast cancer prediction ability (AUC = 0.63 versus 0.56). Finally, a preliminary study on the dual-tree complex wavelet transform showed that wavelet features alone may have value in risk assessment [75].

In an attempt to identify highly discriminative texture descriptors from multiple feature groups and develop optimal combinations that maximize the case-control classification performance, research groups have also explored comprehensive sets of multi-parametric features reflecting various aspects of mammographic texture [51, 53, 56, 66, 77, 78]. Following an evaluation of more than 1000 co-occurrence, run-length, Laws, wavelet, and Fourier features in prior film mammograms of 246 cases and 522 controls, Manduca et al. [66] identified individual features which, when estimated at a coarse scale of a single ROI covering the entire breast, provided strong prediction for future breast cancer (odds ratio per 1 SD = 1.36–1.50, AUC = 0.61–0.62). In another retrospective study with 864 cancer cases and 418 controls, a three-step variable selection process

### Table 3
Breast cancer prediction capacity of automated characterization of the parenchymal patterns (Continued)

| Study          | Institution                        | Cases | Controls | Matched subgroup | Model | AUC | 95% CI       |
|----------------|------------------------------------|-------|----------|------------------|-------|-----|-------------|
| Li et al. (2004) [54] | University of Chicago, University of Pennsylvania | 30    | 60       | Yes              | LDA   | 0.69 | 0.62, 0.75 |
| Li et al. (2005) [81] | University of Chicago              | 30    | 60       | Yes              | ROCA  | 0.66 | 0.58, 0.75 |
| Li et al. (2005) [81] | University of Chicago              | 30    | 60       | Yes              | ROCA  | 0.67 | 0.58, 0.75 |
| Li et al. (2007) [82] | University of Chicago              | 30    | 60       | Yes              | ROCA  | 0.74 | 0.63, 0.85 |
| Li et al. (2007) [82] | University of Chicago              | 30    | 60       | Yes              | ROCA  | 0.77 | 0.67, 0.87 |
| Li et al. (2008) [83] | University of Chicago              | 30    | 60       | No               | ROCA  | 0.90 |             |
| Li et al. (2008) [83] | University of Chicago              | 30    | 60       | Yes              | ROCA  | 0.89 |             |
| Li et al. (2012) [84] | University of Chicago              | 53    | 328      | No               | BANN  | 0.82 |             |
| Li et al. (2012) [84] | University of Chicago              | 34    | 136      | Yes              | BANN  | 0.81 |             |
| Li et al. (2014) [71] | University of Chicago              | 34    | 136      | Yes              | BANN  | 0.81 |             |
| Gierach et al. (2014) [85] | University of Chicago          | 137   | 100      | No               | BANN  | 0.68 | 0.59, 0.72 |
| Gierach et al. (2014) [85] | University of Chicago, NC-I-NIH, Washington Radiology Associates, Genetech, USUHS, UCL, WRNMC, Westat Inc. | 126   | 89       | Yes              | BANN  | 0.71 | 0.55, 0.72 |

Area under the ROC curve (AUC) achieved by risk assessment models when fed with mammographic texture and/or density measures. IMPRS International Max Planck Research School for Optics and Imaging, Moffitt Moffitt Cancer Center and Research Institute, NCI-NIH National Cancer Institute, National Institutes of Health, RadboudUMC Radboud University Nijmegen Medical Centre, TTUHS Texas Tech University Health Sciences, UCL University College London, UCLA University of California at Los Angeles, USUHS Uniformed Services University of the Health Sciences, WRNMC Walter Reed National Military Medical Center
Assessing the risk of carrying a high-risk gene mutation

The potential of mammographic texture in breast cancer risk assessment has also been demonstrated in studies with BRCA1/2 carriers, where texture features from a single 25.6 × 25.6 mm² retro-areolar ROI in CC mammographic views were shown to predict a woman’s risk of carrying this high-risk genetic mutation. The first study addressing this topic extracted a comprehensive feature set of grey-level intensity statistics, co-occurrence features, and multi-scale texture measures based on Fourier transform analysis [80]. In film mammograms of 30 BRCA1/2 carriers and 142 low-risk women, most features demonstrated high individual discriminatory capacity (AUC > 0.68), while the collective performance of the features that were deemed significant in multivariable models raised AUC values of 0.91 and 0.92 in the entire database and in an age-matched subgroup, respectively [55]. Using the same image dataset, the authors also showed a promising individual classification performance for structural measures such as edge frequency (AUC = 0.78) [81], for different implementations of the fractal dimension (AUC = 0.74–0.93) [81, 82], and for power law spectral analysis (AUC = 0.90) [83].

These results were recently replicated and validated in datasets with digital mammograms [71, 84] and larger numbers of high-risk women [71, 84, 85]. A similar design of texture analysis in the retroareolar breast region combined with a Bayesian Artificial Neural Network (BANN) for the classification task was applied to 1) film mammograms of 137 mutation carriers and 100 low-risk women [85], and 2) digital mammograms of 53 mutation carriers, 75 women with unilateral cancer, and 328 low-risk women [71, 84]. The first analysis conferred a two-fold increase in the odds of predicting BRCA1/2 mutation status, and an AUC of 0.68 for texture features alone and 0.72 for the features plus breast PD [85]. In the second analysis, AUC values of 0.82 and 0.73 were obtained between mutation carriers and low-risk women, and between unilateral cancer and low-risk women, respectively [84]; these evaluation results were also retained in age-matched subgroup analysis (0.81 and 0.70, respectively) [84] without any significant improvement from the inclusion of breast PD (0.81 and 0.68, respectively) [71].

Beyond established risk factors in breast cancer risk assessment

A comparison of the evaluation results published to date (Table 3), focusing primarily on cross-validated experiments, suggests that more comprehensive sets of multi-parametric texture features [51, 53, 54, 56, 71, 77, 84] may be more effective in predicting breast cancer than a single feature group. However, the literature lacks extensive comparative studies on the same datasets and generalized conclusions should, therefore, be limited. While the implementation of texture analysis, including both the location and size of ROIs [51, 52, 54] and the specific texture measures, appears to have an effect on texture classification performance, all studies have consistently shown the highly promising, independent role of automated texture analysis in breast cancer risk assessment. Specifically, parenchymal texture descriptors have demonstrated a strong cross-validated ability in
predicting both risk for breast cancer \((0.58 \leq \text{AUC} \leq 0.85)\) and BRCA1/2 mutation status \((0.53 \leq \text{AUC} \leq 0.93)\). Moreover, texture performance has been shown to be either comparable or significantly higher than the performance of breast PD \((0.51 \leq \text{AUC} \leq 0.62\) and \(0.53 \leq \text{AUC} \leq 0.59\), respectively), as reported in studies where texture and density measures were comparatively evaluated on the same datasets \([51, 56, 61, 66, 71, 73–75, 85]\).

In addition, a number of related findings suggest that texture analysis is able to provide complementary information about a woman’s risk of developing breast cancer which cannot be captured by breast PD and other established risk factors. Texture descriptors have been weakly or moderately correlated with breast PD \([55, 61, 64, 71–73, 85–87]\), and weakly correlated with risk factors as reflected in the Gail and Claus risk scores \([80, 86]\). In addition, texture descriptors deemed as strong predictors of breast cancer retained significance when breast PD, age, BMI, family history of breast cancer, parity, age at first term pregnancy, number of previous breast biopsies, menopause, and hormonal use, all shown to be associated with breast cancer risk, were simultaneously considered in classification models (Table 3). Finally, with age-matched datasets or model adjustments for age, most studies evaluating the capacity of parenchymal texture features in risk assessment have ruled out possible confounding due to differences in age, a major breast cancer risk factor, thereby showing a strong potential for computerized texture descriptors in augmenting breast cancer risk assessment.

**Future directions**

Moving forward, experiments evaluating the relative performance of different implementations of texture analysis, using the same evaluation methodology (i.e., dataset and classification model), are necessary to develop more robust and reproducible quantitative mammographic phenotypes of breast cancer risk. Future studies to test the incremental value added by computerized textural measures in predicting breast cancer will require: (a) the design of large age-matched datasets; (b) the selection of an effective classification model, where different previously used models (Table 3) could be comparatively examined; (c) model adjustments to rule out possible confounding due to differences in major risk factors; and (d) validation of the classification performance in independent datasets.

In an attempt to add an anatomical meaning in texture analysis which may also give additional discrimination power to feature classification, increasing attention is currently given to the incorporation of breast anatomy in texture analysis pipelines. Brandt et al. \([74]\) first introduced an anatomically oriented breast coordinate system which allows for anatomical correspondences across mammograms of the same woman or different women. In preliminary analyses using the proposed coordinate system, the authors have demonstrated that anatomy-driven Gaussian derivative features are able to (a) effectively separate cancer cases and controls \([74]\), (b) quantify the effect of hormone replacement therapy as a change in the breast parenchymal patterns \([88]\), and (c) demonstrate specific regions of the breast parenchyma where breast cancer risk is mainly expressed \([89]\). More recently, Gastounioti et al. \([90–92]\) showed that the discriminatory capacity of texture descriptors is further enhanced by an anatomy-driven polar grid for anatomical breast sampling and a breast-anatomy-weighted texture signature which considers the spatial position and the underlying tissue composition of individual ROIs to summarize the parenchymal texture properties of the breast.

Another emerging technology is deep learning \([93]\), which may prove a valuable addition in texture analysis for breast cancer risk assessment \([94–96]\). Deep learning involves automated learning, from raw image data, of hierarchical representations useful for pattern detection and classification, in a supervised mode via neural networks with multiple hidden layers or in an unsupervised mode via autoencoders. The few available studies which have applied deep learning in the particular field show a promising role in risk scoring \((\text{AUC} = 0.61–0.65)\) \([94, 96]\). Further, preliminary comparisons against two previously presented methodologies with hand-crafted texture features \([56, 64]\) suggest that it may be better to “let the data speak” instead of modeling prior assumptions \([94]\). Additional experimentation with deep learning, as well as future comparisons with the state-of-the-art texture analysis techniques, is warranted to better explore the potential of this novel technology.

Digital breast tomosynthesis (DBT), an emerging x-ray technology \([97]\) in which quasi three-dimensional (3D) images are reconstructed from a limited number of low-dose x-ray source projections \([98, 99]\), is increasingly being implemented clinically due to improvements in sensitivity and specificity compared to imaging with digital mammography alone \([100]\). By imaging the breast in 3D, DBT alleviates the effect of tissue superimposition, offering superior tissue visualization, which in turn may allow for better characterization of the breast parenchyma compared to two-dimensional mammography \([86, 87]\). The extension of the parenchymal texture analysis descriptors for volumetric texture analysis in DBT is, therefore, an important future challenge towards developing superior texture features which can optimize image-driven breast cancer risk assessment.

Another challenging future step which would establish the predictive value of texture analysis is the validation
of parenchymal texture measures in prospectively collected data. Large-scale studies involving multiple screening centers, imaging machines, and image acquisition settings are also of major importance towards validating their predictive capacity and robustness to heterogeneous image data [101]. Furthermore, the literature lacks large-scale longitudinal studies monitoring longitudinal changes in automated parenchymal texture descriptors over successive mammograms, which could elucidate the mechanisms of breast cancer development [11] and the causal relations between the texture risk scoring and breast cancer [102, 103]. Finally, crucial questions to be addressed in such rich datasets are the causes of inter-woman variation in mammographic parenchymal patterns [104, 105] and in the relation of texture risk markers to the subsequent location and grading of tumors, disease mortality, and treatment effects [20, 21].

The valuable risk markers provided by parenchymal texture analysis could also leverage the relatively new, yet promising, paradigms of radiomics [106] and radiogenomics [107] for breast cancer, aiming to convert breast images into comprehensive measurable data and to delve into the interaction between these data and genetic variants. These novel approaches may pave the way to revealing correlations with the genomic diversity present in breast cancer, understanding how biological processes are reflected in quantitative breast imaging phenotypes, and defining novel clinical biomarkers or biological surrogates [108–111], thus improving personalized breast cancer screening, monitoring, and treatment selection.

Conclusions
Automated breast parenchymal texture analysis has the potential to elucidate imaging phenotypes of breast cancer risk, which is valuable in accelerating the translation of individualized risk stratification into routine breast cancer screening and prevention strategies. Future work addressing technical challenges in this field and large prospective studies are expected to further enhance and establish the predictive value of parenchymal texture measures for inclusion in breast cancer risk assessment models in clinical practice.

Abbreviations
3D: Three-dimensional; AUC: Area under the ROC curve; CC: Cranio-caudal; DBT: Digital breast tomosynthesis; ER−: Estrogen receptor negative; ER+: Estrogen receptor positive; GLCM: Grey-level co-occurrence matrix; MLO: Medio-lateral oblique; MTR: Mammographic texture resemblance; PD: Percent density; ROI: Region of interest; RR: Relative risk

Funding
The authors wish to acknowledge support by the National Cancer Institute at the National Institutes of Health via a research grant award (5R01CA163313), the Population-based Research Optimizing Screening through Personalized Regimens (PROSPR) Network (U54CA163313), and a Resource-Related Research Project–Cooperative Agreement (1U24CA189523).

Authors’ contributions
AG performed the literature research and drafted the manuscript. EFC participated in the design of the review and drafting of the manuscript. DK coordinated the review, participated in the design of the review, and drafting of the manuscript. All authors have read and approved the final manuscript and agree to be accountable for all aspects of the work.

Competing interests
The authors declare that they have no competing interests.

Published online: 20 September 2016

References
1. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. GLOBOCAN 2012 v1.0, cancer incidence and mortality worldwide: IARC CancerBase No. 11 Lyon, International Agency for Research on Cancer. France, 2013. http://globocan.iarc.fr. Accessed 8 Mar 2016.
2. Cancer facts and figures 2016 Atlanta, GA: American Cancer Society; 2016. http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2016/. Accessed 8 Mar 2016.
3. Hall P, Easton D. Breast cancer screening: time to target women at risk. Br J Cancer. 2013;108(1):2020–4.
4. Howell A, Astley S, Warwick J, Stavrinos P, Sahin S, Ingham S, et al. Prevention of breast cancer in the context of a national screening programme. J Int Med. 2012;271(4):321–30.
5. Amir E, Freedman OC, Seruga B, Evans DG. Assessing women at high risk of breast cancer: a review of risk assessment models. J Natl Cancer Inst. 2010;102(10):680–91.
6. Gail MH, Mai PL. Comparing breast cancer risk assessment models. J Natl Cancer Inst. 2010;102(10):665–8.
7. Omega T, Beaber EF, Sprague BL, Barlow WE, Haas JS, Tosteson AN, et al. Breast cancer screening in an era of personalized regimens: a conceptual model and National Cancer Institute initiative for risk-based and preference-based approaches at a population level. Cancer. 2014;120(19):2955–64.
8. McDonald ES, Clark AS, Tchou J, Zhang P, Freedman GM. Clinical diagnosis and management of breast cancer. J Nucl Med. 2016;57(Supplement 1):S5–16S.
9. Chen J-H, Gulien G, Su M-Y. Imaging breast density: established and emerging modalities. Transl Oncol. 2015;8(6):435–45.
10. Ng K-H, Lau S. Vision 20/20: Mammographic breast density and its clinical applications. Med Phys. 2015;42(12):7059–77.
11. Sherratt MJ, McConnell JC, Streuli CH. Raised mammographic density: causative mechanisms and biological consequences. Breast Cancer Res. 2016;18(1):1.
12. McCormack VA, de Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev. 2006;15(6):1159–69.
13. Brentnall AR, Harkness EF, Astley SM, Donnelly LS, Stavrinos P, Sampson S, et al. Mammographic density adds accuracy to both the Tyrer-Cuijvert and Gail breast cancer risk models in a prospective UK screening cohort. Breast Cancer Res. 2015;17(1):1–10.
14. Tice JA, Cummings SR, Smith-Blindman R, Ichikawa L, Barlow WE, Kerlikowske K. Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model. Ann Intern Med. 2008;148(5):337–47.
15. Are You Dense Advocacy. D.E.N.S.E. State Efforts. http://areyoudenseadvocacy.org. Accessed 1 June 2016.
16. Abdellah M, Tsuruda K, Lightfoot CB, Payne JI, Caines J, Iles SE. Utility of relative and absolute measures of mammographic density versus clinical risk factors in evaluating breast cancer risk at time of screening mammography. Br J Radiol. 2016;89(1059):20150522.
17. Tan M, Zheng B, Ramalingam S, Gur D. Prediction of near-term breast cancer risk based on bilateral mammographic feature asymmetry. Acad Radiol. 2013;20(12):1542–50.
18. Tan M, Zheng B, Leader J, Gur D. Association between changes in mammographic image features and risk for near-term breast cancer development. IEEE Trans Med Imaging. 2016;35(7):1719–28.
