



# Automated rating of background parenchymal enhancement in MRI of extremely dense breasts without compromising the association with breast cancer in the DENSE trial

Hui Wang<sup>a</sup>, Bas H.M. van der Velden<sup>a</sup>, Erik Verburg<sup>a</sup>, Marije F. Bakker<sup>b</sup>, Ruud M. Pijnappel<sup>c</sup>, Wouter B. Veldhuis<sup>c</sup>, Carla H. van Gils<sup>b</sup>, Kenneth G.A. Gilhuijs<sup>a,\*</sup>

<sup>a</sup> Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands

<sup>b</sup> Julius Center for Health Sciences and Primary Care, Utrecht, The Netherlands

<sup>c</sup> Department of Radiology, University Medical Center Utrecht, The Netherlands

## ARTICLE INFO

### Keywords:

Background Parenchymal Enhancement  
Breast MRI  
Machine learning  
Dense breasts

## ABSTRACT

**Objectives:** Background parenchymal enhancement (BPE) on dynamic contrast-enhanced MRI (DCE-MRI) as rated by radiologists is subject to inter- and intrareader variability. We aim to automate BPE category from DCE-MRI. **Methods:** This study represents a secondary analysis of the Dense Tissue and Early Breast Neoplasm Screening trial. 4553 women with extremely dense breasts who received supplemental breast MRI screening in eight hospitals were included. Minimal, mild, moderate and marked BPE rated by radiologists were used as reference. Fifteen quantitative MRI features of the fibroglandular tissue were extracted to predict BPE using Random Forest, Naïve Bayes, and KNN classifiers. Majority voting was used to combine the predictions. Internal-external validation was used for training and validation. The inverse-variance weighted mean accuracy was used to express mean performance across the eight hospitals. Cox regression was used to verify non inferiority of the association between automated rating and breast cancer occurrence compared to the association for manual rating. **Results:** The accuracy of majority voting ranged between 0.56 and 0.84 across the eight hospitals. The weighted mean prediction accuracy for the four BPE categories was 0.76. The hazard ratio (HR) of BPE for breast cancer occurrence was comparable between automated rating and manual rating (HR = 2.12 versus HR = 1.97,  $P = 0.65$  for mild/moderate/marked BPE relative to minimal BPE). **Conclusion:** It is feasible to rate BPE automatically in DCE-MRI of women with extremely dense breasts without compromising the underlying association between BPE and breast cancer occurrence. The accuracy for minimal BPE is superior to that for other BPE categories.

## 1. Introduction

Women aged between 50 and 75 years in the Netherlands are screened for breast cancer every two years using mammography [1]. Mammography is, however, less sensitive to detect breast cancer in women with extremely dense breasts (i.e., women in category d of the Breast Imaging Reporting and Data System (BIRADS) lexicon) [2]. Women with extremely dense breasts have three to six times higher risk of developing breast cancer compared to women with almost entirely fatty breasts (i.e., women in category a of the BIRADS lexicon) [3,4]. A recent randomized controlled trial– Dense Tissue and Early Breast Neoplasm Screening (DENSE) investigated whether MRI has

complementary value for the detection of breast cancer in a mammography-screening population of women aged 50–75 years with extremely dense breasts. Based on the results of this national multi-institutional study, it would be helpful to identify other risk factors in addition to breast density that could be used to personalize breast cancer screening and reduce the workload of radiologists and costs for the community.

Evaluation of background parenchymal enhancement (BPE) has recently gained more attention due to its association with breast-cancer risk. Several studies have shown that the amount of BPE in dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) was predictive of breast cancer [5–7], but whether this applies to women in all

\* Corresponding author at: University Medical Center Utrecht, Q.02.4.45, P.O. Box 85500, 3508 GA Utrecht, The Netherlands.

E-mail address: [K.G.A.Gilhuijs@umcutrecht.nl](mailto:K.G.A.Gilhuijs@umcutrecht.nl) (K.G.A. Gilhuijs).

<https://doi.org/10.1016/j.ejrad.2024.111442>

Received 24 October 2023; Received in revised form 6 February 2024; Accepted 21 March 2024

Available online 24 March 2024

0720-048X/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

risk groups is still under debate [8]. In a group of women at high risk of breast cancer but without a history of the disease, researchers found that those with mild, moderate or marked BPE had a nine times higher risk of developing breast cancer compared with those with minimal BPE [9]. Moreover, BPE was found to be associated with treatment response [10–13] and prognosis [14–16]. These findings suggest that BPE can potentially serve as an important biomarker of risk and prognosis. However, radiologists currently lack objective tools for assessing the category of BPE.

In clinical practice, BPE is evaluated qualitatively by radiologists using the BIRADS categories of minimal, mild, moderate, or marked. This assessment is, however, subject to intra- and inter- reader variability. Kappa values ranged from moderate to almost perfect for intra-reader agreement, and from fair to almost perfect for inter-reader agreement [6,17,18].

The aim of this study is to investigate whether automated rating of BPE category on DCE- MRI of women with extremely dense breasts can be achieved while preserving underlying associations between BPE and breast cancer occurrence.

2. Materials and Methods

2.1. Study population

This is a secondary analysis of data from the DENSE trial [1,19,20]. The DENSE trial has been approved by the Dutch Minister of Health, Welfare and Sport (2011/19 WBO, The Hague, the Netherlands). The study including primary and secondary objectives was waived from ethical review by the local institutional review board based on the Dutch law on population studies. The DENSE trial was designed to compare the effectiveness of breast cancer screening with mammography alone versus mammography and MRI in women aged 50–75 years with extremely dense breasts (i.e., breast density category d at mammography) [21]. The participants of this study were recruited from eight Dutch hospitals without abnormality on mammography at the time of inclusion. The MRI screenings were performed in three consecutive rounds starting in 2011 and concluding with the third-round screening in 2021. Written informed consent was obtained from all women who underwent MRI. In the current study, the first-round screening images only were analyzed. Women were excluded if they had bilateral cancer or had unilateral cancer and had previously undergone contralateral mastectomy. All malignancies were confirmed by histopathology according to the DENSE trial protocol based on available consensus [22,23]. As such, B3 lesions such as lobular neoplasia and atypical ductal hyperplasia were classified as benign.

2.2. MRI acquisition

The study participants underwent MRI screening using either a Philips 3.0 T scanner (Achieva or Ingenia) or a Siemens 3.0 T scanner (Trio, Verio, or Skyra). The MRI screening was conducted in the prone position and included a dynamic series with a precontrast image and 4 or 5 postcontrast images after using a gadolinium-based contrast medium (gadobutrol, Gadovist; Bayer Healthcare, Germany). Contrast agent was injected at a rate of 1 mL/s to a total dose of 0.1 mmol per kilogram of body weight. Fat suppression was optional during DCE-MRI acquisition. The imaging parameters have been described in detail elsewhere [1] and are summarized in Supplemental Table 1.

2.3. Image analysis

The image analysis included deformable image registration [24], breast segmentation [25], FGT segmentation, and feature extraction. The FGT was automatically segmented from the precontrast images using an nnU-Net [26] convolutional neural network in an iterative learning approach [7].

Table 1  
Description of the Parenchyma Features.

Circularity	Ratio between the intersection of the FGT with a sphere of equivalent volume as the FGT and the volume of the FGT itself
Irregularity	Ratio between the surface of a sphere with the same volume as the FGT and the surface of the FGT
Density	Ratio between the volume of FGT and the volume of the whole breast
Top 10 % /50 %/90 % intensity of early contrast uptake	Top 10 %/ 50 %/ 90 % early enhancing intensity of the contralateral healthy breast (women with cancer) or mean value of top 10 %/ 50 %/ 90 % of both breasts (women without cancer)
Top 10 %/ 50 %/ 90 % volume of early enhancing voxels	Volume of the top 10 %/ 50 %/ 90 % early enhancing voxels of the contralateral healthy breast (women with cancer) or mean value of top 10 %/ 50 %/ 90 % of both breasts (women without cancer)
Top 10 % /50 %/ 90 % intensity of late contrast uptake	Top 10 % /50 %/ 90 % late enhancing intensity of the contralateral healthy breast (women with cancer) or mean value of top 10 %/ 50 %/ 90 % of both breasts (women without cancer)
Top 10 % /50 %/ 90 % volume of late enhancing voxels	Volume of the top 10 % /50 %/ 90 % late enhancing voxels of the contralateral healthy breast (women with cancer) or mean value of top 10 % /50 %/ 90 % of both breasts (women without cancer)

FGT: fibroglandular tissue

The image analysis was conducted using Python (version 2.7; Python Software Foundation, Beaverton, OR) and PyTorch (version 1.5).

2.4. Feature extraction

Fifteen quantitative features describing the spatiotemporal characteristics of FGT were extracted, including volumetric density, volumetric morphology (i.e., the three-dimensional shape of the FGT), and enhancement characteristics. These features have been described in detail in a previous study [7] and are summarized in Table 1.

2.5. Manual rating of BPE category

In the DENSE trial, BPE category was evaluated by experienced radiologists using the BIRADS MRI lexicon according to the DENSE trial protocol. The BPE level was assessed on the first postcontrast acquisition on T1 weighted images, at or before 120 s after contrast agent injection. The radiologists rated the BPE category in four categories: minimal, mild, moderate, and marked BPE. When BPE assessment between breasts of the same participant were inconsistent, the breast with the greater amount of BPE was used for the overall assessment. The experience of these radiologists in reading breast MRI ranged between 5 and 23 years at the start of the trial.

2.6. Statistical analysis

Multiple imputation was used to impute missing values. Ten imputation sets were merged based on the mean imputed values [27–29]. Outliers were defined based on Tukey’s method, i.e., 1.5 interquartile range below the 25th percentile or 1.5 interquartile range above the 75th percentile. Outliers were winsorized to the nearest whisker after which the feature values were normalized between 0 and 1.

Internal-external validation was used for training and validation. In short, seven of eight hospitals were used to train the machine-learning method using 5-fold cross validation, and the remaining hospital was then used to perform external validation. This process was repeated eight times, with each hospital serving as the external validation hospital once. The results were merged. In each internal validation round, bootstrapping was performed with 100 bootstrap cycles to assess the

uncertainty. In each bootstrap cycle, three classifiers (i.e., Random Forest, Naïve Bayes, and KNN) were constructed separately to predict BPE category. Majority voting [30–32] was used to combine the prediction outcome of these three classifiers.

The accuracy of the majority voting on the external validation set was used as metric to evaluate prediction performance in each bootstrap cycle. Inverse-variance weighted mean accuracy was used to merge the external validation performance of these eight hospitals [33].

Cox regression was used to verify if the association between automatically rated BPE and breast cancer occurrence – expressed as Hazard Ratio – was comparable to that observed between manual rating of BPE and breast cancer occurrence. All malignancies were confirmed by histopathology conform the DENSE trial protocol based on available consensus [22,23].

The statistical analyses were performed in R (version 4.1.2; R Foundation for Statistical Computing, Vienna, Austria), the classifier construction process was performed using the Caret package (version 6.0–92).

### 3. Results

#### 3.1. Study population

In total, the data from 4783 women in the first round of the DENSE trial were available from the eight hospitals. Women with incomplete digital MRI data or who met the exclusion criteria were excluded (Fig. 1). The mean age of the remaining 4553 study participants with extremely dense breasts was 56 years.

#### 3.2. Feature extraction

For 120 of the 4553 women, BPE category (assessed by radiologists) was not available in the current study. For these 120 women, the scores were imputed. BPE categories minimal (75 %) and mild (15 %) were scored most often (Table 2). In hospital 6, women less often received BPE category 'minimal' (55 %) compared to women from the other seven hospitals (67 % to 80 %). A higher fraction of women in hospital 6 had a BPE category of 'mild' (30 %) compared to the fraction in other seven hospitals (10 % to 17 %) (Fig. 2).

#### 3.3. Prediction of BPE category

The accuracy of the random forest, Naïve bayes, and k-NN classifiers to predict the BPE score in the eight hospitals ranged between 0.30 and 0.84 (Table 3). The accuracy of the majority voting ranged between 0.56 and 0.84 (Table 3). No obvious difference in accuracy was observed between centers, despite the use of scanners from different

**Table 2**

Distribution of clinical characteristics of the 4553 women from the DENSE trial.

Age(mean ± SD), years	55.7 ± 6
BPE category	
Minimal	3436(75 %)
Mild	663(15 %)
Moderate	273(6 %)
Marked	181(4 %)

Numbers represent frequency unless stated otherwise; SD: standard deviation; BPE: background parenchymal enhancement

manufactures. The cross-validated prediction accuracy of BPE based on majority voting was, however, somewhat lower in hospital 6 (0.56) compared to that in the other hospitals (0.68 to 0.84). The sensitivity of the majority voting method was highest for predicting the presence of 'minimal' BPE (0.95) compared to other BPE categories (0.13 to 0.36) (Table 4). This suggests that machine learning was most effective at accurately identifying minimal BPE among study participants. Conversely, the specificity of the majority voting was lowest for 'minimal' BPE (0.46) compared to the other BPE categories (0.95 to 0.97) (Table 4). The overall weighted mean accuracy established from the eight hospitals was 0.76.

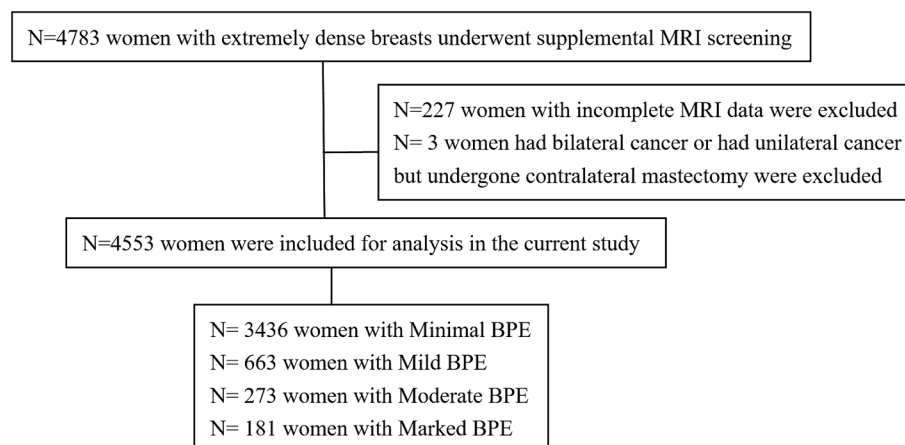
#### 3.4. Association with breast cancer occurrence

The median follow-up time was 47 months [95 %CI: 16–55 months]. Among the included women, 122 of 4553 (3 %) were diagnosed with breast cancer, and the mean time to cancer detection was 12.4 months. Regardless whether BPE was rated manually (HR = 1.97,  $P < 0.001$ ) or automatically (HR = 2.12,  $P < 0.001$ ), categories higher than minimal (i.e., mild/moderate/marked) were associated with higher breast cancer occurrence than category minimal (Table 5).

Additionally, the hazard ratio of BPE for breast cancer occurrence was comparable between automated rating and manual rating (HR = 2.12 versus HR = 1.97,  $P = 0.65$  for mild/moderate/marked BPE relative to minimal BPE).

### 4. Discussion

This study aimed to determine whether it is feasible to use machine learning to automatically evaluate BPE category on dynamic contrast-enhanced MRI (DCE-MRI) of women with extremely dense breasts without compromising associations with breast cancer occurrence. A combination of Random Forests, Naïve Bayes, and KNN classifiers with majority voting was used to develop a machine-learning model that



**Fig. 1.** Flowchart of included women. BPE: Background parenchymal enhancement (BPE was assessed by radiologists).

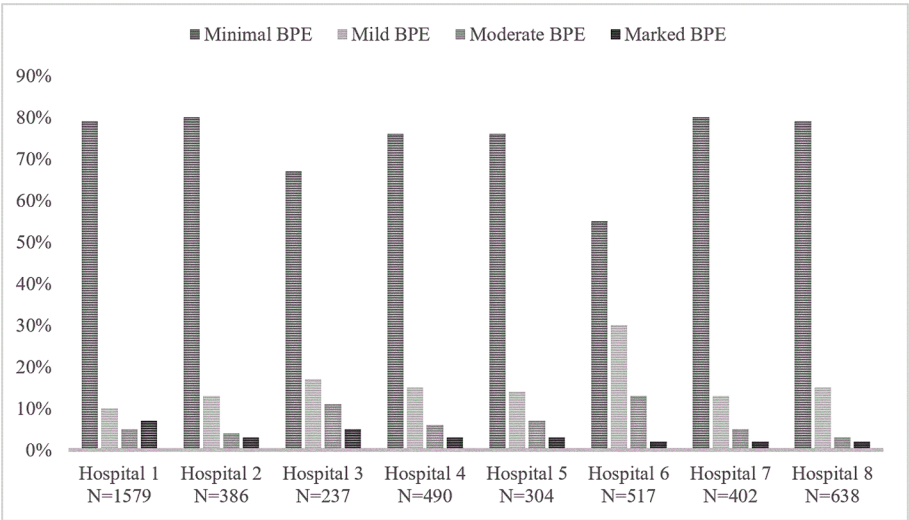


Fig. 2. Percentage distribution of BPE categories in the eight hospitals participating in the DENSE trial.

Table 3  
External validation accuracy of 100 times bootstrapping per hospital per classifier.

Hospital ID for external validation	External validation size	Training size	Random forest	Naïve bayes	KNN	Majority Voting	Weighted accuracy
Hospital 1	1579	2974	0.82 [0.81,0.82]	0.79 [0.79,0.79]	0.81 [0.81,0.81]	0.82 [0.81,0.82]	0.76
Hospital 2	386	4167	0.82 [0.81,0.82]	0.57 [0.57,0.57]	0.80 [0.80,0.80]	0.80 [0.80,0.81]	
Hospital 3	237	4316	0.71 [0.70,0.73]	0.72 [0.72,0.72]	0.72 [0.72,0.72]	0.71 [0.71,0.73]	
Hospital 4	490	4063	0.80 [0.79,0.81]	0.76 [0.76,0.76]	0.82 [0.82,0.82]	0.8 [0.8,0.81]	
Hospital 5	304	4249	0.80 [0.79,0.81]	0.75 [0.75,0.75]	0.79 [0.79,0.79]	0.79 [0.78,0.79]	
Hospital 6	517	4036	0.56 [0.56,0.56]	0.57 [0.57,0.57]	0.57 [0.57,0.57]	0.56 [0.56,0.57]	
Hospital 7	402	4151	0.83 [0.82,0.84]	0.83 [0.83,0.83]	0.84 [0.84,0.84]	0.84 [0.83,0.85]	
Hospital 8	638	3915	0.71 [0.70,0.72]	0.30 [0.30,0.30]	0.69 [0.69,0.69]	0.68 [0.67,0.69]	

Numbers in column 4 to 7 represent mean accuracy [95 % confidence interval]

Table 4  
Prediction performance in the four BPE categories.

BPE category	Minimal	Mild	Moderate	Marked
Sensitivity	0.95 [0.95, 0.95]	0.16 [0.16, 0.17]	0.13 [0.12, 0.14]	0.36 [0.34, 0.37]
Specificity	0.46 [0.45, 0.47]	0.95 [0.94, 0.95]	0.96 [0.96, 0.97]	0.97 [0.97, 0.97]

Numbers in column 2 ~ 5 represent median [95 % confidence interval]

could predict BPE category with a cross-validated prediction accuracy of 0.76. The results indicate that it is feasible to automate the evaluation of BPE category in women with extremely dense breasts, although the accuracy for minimal BPE is superior to that for other BPE categories. The inter-observer variability in BPE has been well documented [17], i. e., different radiologists will not always agree on BPE rating. However, our results show that the underlying association between BPE and breast cancer occurrence – expressed as hazard ratio – was not affected and was comparable with that between manually rated BPE in the DENSE trial and breast cancer occurrence.

The cross-validated prediction accuracy ranged between 0.56 and 0.84 across the eight hospitals. The lowest accuracy was observed in

Table 5  
Association between BPE category and occurrence of breast cancer for manual scoring and scoring by machine learning.

Characteristics	Hazard Ratio (HR)	95 %CI	P value of HR	P value of comparing HRs
<b>BPE rated by radiologist</b>				
Minimal	Reference	–	–	
Mild/Moderate/Marked	1.97	[1.37, 2.84]	<0.001	
<b>BPE predicted by ML</b>				
Minimal	Reference	–	–	
Mild/Moderate/Marked	2.12	[1.43, 3.16]	<0.001	

BPE: Background parenchymal enhancement, CI: confidence interval, ML: Machine Learning

hospital 6 (0.56). This may be attributable to the fact that radiologists in hospital 6 typically assigned higher scores to BPE. Consequently, hospital 6 had the lowest percentage of women in the minimal BPE category (55 %), while the sensitivity of the machine-learning model for this category was the highest. The percentage of women in the minimal BPE category ranged between 67 % and 80 % in the other hospitals.



In clinical practice, BPE is rated by radiologists according to the BIRADS lexicon, and is subject to inter- and intrareader variation [17], which may have hindered its role as predictive imaging biomarker of breast cancer. Quantitative evaluation of BPE can overcome this disadvantage. A previous study assessed quantitative parenchymal features at baseline DCE MRI and their association with breast cancer occurrence [7], using manual rating of BPE to confirm that such quantitative analysis is independently associated with breast cancer occurrence. The current study focuses explicitly on automated rating of BPE to help further reduce inter- and intra-observer variability. It is yet unknown whether longitudinal updates of BPE will improve assessment of breast-cancer risk, and it is subject of future research.

Several other studies reported on automated prediction of BPE category from DCE MRI, but the study populations and methodology vary considerably.

A recent study from Nam et al. investigated 794 patients with breast cancer who underwent preoperative breast MRI from 2014 to 2017. Deep learning was applied to automatically assess BPE category. The overall classification accuracy among the four BPE categories was 0.67 [34]. In contrast, the current study focuses only on an unselected series of asymptomatic women with extremely dense breasts. Although MRI is typically indicated for screening of women at high life-time risk of breast cancer, women in the DENSE trial were recruited from the national mammography screening program instead, having no apparent risk factors other than having extremely dense breasts. Hence, the number of women at high life-time risk in the DENSE trial is expected to be very small.

Borkowski et al. trained a deep convolutional neural network for the classification of BPE in 149 patients who underwent breast MRI from September 2013 to October 2015, yielding mean accuracy of 0.75 in external validation [35]. Neither the study by Borkowski et al. nor the study by Nam et al. specifically focus on women with extremely dense breasts. Hence, it is difficult to assess the potential of these findings in a risk stratification tool for this specific screening population.

Sarah et al. recently reported on 3705 high-risk women (i.e. women with > 20 % lifetime risk of breast cancer) using 5224 breast MRI examinations [36]. Two deep learning models were constructed to distinguish between low (i.e., mild or minimal) and high (i.e., marked or moderate) BPE. Although the study uses a relatively large sample size, it also does not focus explicitly on women with extremely dense breasts.

## 5. Limitations

The current study has some limitations. Firstly, although the model was validated in different hospitals using internal-external validation, the results have not yet been validated in other external datasets, e.g., those from other countries. Wider validation of the model is recommended prior to clinical application. Secondly, the number of women in each BPE category was not balanced. Although this reflects the natural distribution of BPE in the population of Dutch women with extremely dense breasts, it may have impact on the machine learning classifier, especially in women with moderate and marked BPE. In the future, the model could be investigated for its efficacy in other populations of women at risk. To the best of our knowledge, no studies have reported on the impact of the precise timing of the first postcontrast series on BPE. Particularly in the context of BPE for risk stratification, further investigation into this aspect may be warranted.

## 6. Conclusion

It is feasible to rate BPE category automatically in contrast-enhanced MRI of women with extremely dense breasts without compromising the association between BPE and breast cancer occurrence, although the accuracy for minimal BPE is superior to that of other BPE categories.

## Funding

This study is financially supported by the Dutch Cancer Society (KWF Kankerbestrijding), grant number UU-2014-7151, and used data acquired during the DENSE trial. The DENSE trial was supported by the regional screening organizations, Volpara Solutions, the Dutch Expert Centre for Screening, and the National Institute for Public Health and the Environment. The DENSE trial is financially supported by the University Medical Center Utrecht (project number: UMCU DENSE), the Netherlands Organization for Health Research and Development (ZonMw, project numbers ZONMW-200320002-UMCU and ZonMW Preventie 50-53125-98-014), the Dutch Cancer Society (KWF Kankerbestrijding, project numbers c), the Dutch Pink Ribbon/A Sister's Hope (project number: Pink Ribbon-10074), Bayer AG Pharmaceuticals, Radiology (project number BSP-DENSE), and Stichting Kankerpreventie Midden-West.

## Informed Consent

Written informed consent was obtained from all women who underwent MRI.

## Ethical Approval

This is a secondary analysis of data from the DENSE trial. The DENSE trial has been approved by the Dutch Minister of Health, Welfare and Sport (2011/19 WBO, The Hague, the Netherlands). The study including primary and secondary objectives was waived from ethical review by the local institutional review board based on the Dutch law on population studies.

## CRedit authorship contribution statement

**Hui Wang:** Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Bas H.M.van der Velden:** . **Erik Verburg:** Writing – review & editing, Software, Investigation, Data curation, Conceptualization. **Marrie F. Bakker:** Writing – review & editing, Data curation. **Ruud M. Pijnappel:** Writing – review & editing, Methodology. **Wouter B. Veldhuis:** Writing – review & editing, Investigation, Conceptualization. **Carla H. van Gils:** Writing – review & editing, Resources, Methodology, Investigation, Funding acquisition. **Kenneth G. A. Gilhuijs:** Writing – review & editing, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

## Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: **H.W.** No relevant relationships. **B.H.M.v.d.V.** No relevant relationships. **E.V.** No relevant relationships. **M.F.B.** No relevant relationships. **R.M.P.** No relevant relationships. **W.B.V.** No relevant relationships. **C. H.V.G.** Consultant for Bayer from February 2020 to June 2021; speaker fee and compensation for travel from the Swedish Society of Breast Radiology and Canadian Association of Radiology; Director of Dutch Cancer Society since June 1, 2021. **K.G.A.G.** No relevant relationships.

## Acknowledgements

The DENSE trial study group is acknowledged for making the trial data available for this study. The authors acknowledge the study participants for their contributions. Volpara Health provided software for measuring breast density in the DENSE trial. Authors also thank the registration team of the Netherlands Comprehensive Cancer Organisation (IKNL) for the collection of data for the Netherlands Cancer

## Registry.

## References

- [1] M.J. Emaus, M.F. Bakker, P.H.M. Peeters, et al., MR imaging as an additional screening modality for the detection of breast cancer in women aged 50–75 Years with extremely dense breasts: the DENSE trial study design, *Radiology* 277 (2015) 527–537.
- [2] J.O.P. Wanders, K. Holland, W.B. Veldhuis, et al., Volumetric breast density affects performance of digital screening mammography, *Breast Cancer Res Treat* 162 (2017) 95–103.
- [3] N.F. Boyd, H. Guo, L.J. Martin, Mammographic density and the risk and detection of breast cancer, *N Engl J Med* 356 (2007) 227–236.
- [4] N.F. Boyd, L.J. Martin, M.J. Yaffe, et al., Mammographic density and breast cancer risk: current understanding and future prospects, *Breast Cancer Res* 13 (2011) 1–12.
- [5] V.A. Arasu, D.L. Miglioretti, B.L. Sprague, et al., Population-based assessment of the association between magnetic resonance imaging background parenchymal enhancement and future primary breast cancer risk, *J Clin Oncol* 37 (2019) 954–963.
- [6] V. King, J.D. Brooks, J.L. Bernstein, et al., Background Parenchymal enhancement at breast MR imaging and breast cancer risk, *Radiology* 260 (2011) 50–60.
- [7] H. Wang, V.B. Van Der, E. Verburg, et al., Assessing quantitative Parenchymal features at baseline dynamic contrast-enhanced MRI and cancer occurrence in women with extremely dense breasts, *Radiology* 308 (2023) e222841.
- [8] C.M. Thompson, I. Mallawaarachchi, D.K. Dwivedi, et al., The association of background parenchymal enhancement at breast mri with breast cancer: a systematic review and meta-analysis, *Radiology* 292 (2019) 552–561.
- [9] B.N. Dontchos, H. Rahbar, S.C. Partridge, et al., Are qualitative assessments of background Parenchymal enhancement, amount of Fibroglandular tissue on MR images, and mammographic density associated with breast cancer risk? *Radiology* 276 (2015) 371–380.
- [10] S.Y. Kim, N. Cho, Y. Choi, et al., Factors affecting pathologic complete response following neoadjuvant chemotherapy in breast cancer: development and validation of a predictive nomogram, *Radiology* 299 (2021) 290–300.
- [11] S.Y. Kim, N. Cho, S.U. Shin, et al., Contrast-enhanced MRI after neoadjuvant chemotherapy of breast cancer: lesion-to-background parenchymal signal enhancement ratio for discriminating pathological complete response from minimal residual tumour, *Eur Radiol* 28 (2018) 2986–2995.
- [12] C. You, W. Peng, W. Zhi, et al., Association between background Parenchymal enhancement and pathologic complete remission throughout the neoadjuvant chemotherapy in breast cancer patients, *Transl Oncol* 10 (2017) 786–792.
- [13] T. Hilal, M. Covington, H.E. Kosiorek, et al., Breast MRI phenotype and background parenchymal enhancement may predict tumor response to neoadjuvant endocrine therapy, *Breast J* 24 (2018) 1010–1014.
- [14] M. Zhang, M. Sadinski, D. Haddad, et al., Background Parenchymal enhancement on breast MRI as a prognostic surrogate: Correlation with breast cancer oncotype dx score, *Front Oncol* 10 (2021) 1–7.
- [15] S. Moliere, I. Oddou, V. Noblet, et al., Quantitative background parenchymal enhancement to predict recurrence after neoadjuvant chemotherapy for breast cancer, *Sci Rep* 9 (2019) 1–8.
- [16] M.S. Bae, J.M. Chang, N. Cho, et al., Association of preoperative breast MRI features with locoregional recurrence after breast conservation therapy, *Acta Radiol* 59 (2018) 409–417.
- [17] A. Melsaether, M. McDermott, D. Gupta, et al., Inter- and intrareader agreement for categorization of background parenchymal enhancement at baseline and after training, *Am J Roentgenol* 203 (2014) 209–215.
- [18] A. Tagliafico, B. Bignotti, G. Tagliafico, et al., Quantitative evaluation of background parenchymal enhancement (BPE) on breast MRI. a feasibility study with a semi-automatic and automatic software compared to observer-based scores, *Br J Radiol* 88 (2015).
- [19] S.G.A. Veenhuizen, S.V. de Lange, M.F. Bakker, Supplemental breast MRI for women with extremely dense breasts: results of the second screening round of the DENSE trial, *Radiology* 299 (2021) 278–286.
- [20] M.F. Bakker, S.V. De Lange, R.M. Pijnappel, et al., Supplemental MRI screening for women with extremely dense breast tissue, *N Engl J Med* 381 (2019) 2091–2102.
- [21] Spak DA, Plaxco JS, Santiago L, et al. BI-RADS fifth edition : A summary of changes. *Diagn Interv Imaging* 2017; 98:179–190.
- [22] Rageth CJ, O'Flynn EAM, Pinker K, et al. Second International Consensus Conference on lesions of uncertain malignant potential in the breast (B3 lesions). *Breast Cancer Res Treat* 2019; 174:279–296.
- [23] E. Verburg, C.H. van Gils, M.F. Bakker, et al., Computer-aided diagnosis in Multiparametric magnetic resonance imaging screening of women with extremely dense breasts to reduce false-positive diagnoses, *Invest Radiol* 55 (2020) 438–444.
- [24] S. Klein, M. Staring, K. Murphy, et al., elastix: a toolbox for intensity-based medical image registration, *IEEE Trans Med Imaging* 29 (2010) 196–205.
- [25] B.H.M. van der Velden, I. Dmitriev, C.E. Loo, et al., Association between Parenchymal enhancement of the contralateral breast in dynamic contrast-enhanced MR imaging and outcome of patients with unilateral invasive breast cancer, *Radiology* 276 (2015) 675–685.
- [26] F. Isensee, P.F. Jaeger, S.A.A. Kohl, et al., nnU-net: a self-configuring method for deep learning-based biomedical image segmentation, *Nat Methods* 18 (2021) 203–211.
- [27] R.A. Burns, P. Butterworth, K.M. Kiely, et al., Multiple imputation was an efficient method for harmonizing the mini-mental state examination with missing item-level data, *J Clin Epidemiol* 64 (2011) 787–793.
- [28] Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, Inc.; 1987.
- [29] A. Marshall, D.G. Altman, R.L. Holder, et al., Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines, *BMC Med Res Methodol* 9 (2009) 1–8.
- [30] R. Hastie, T. Kameda, The robust beauty of majority rules in group decisions, *Psychol Rev* 112 (2005) 494–508.
- [31] R.D. Sorkin, C.J. Hays, R. West, Signal-detection analysis of group decision making, *Psychol Rev* 108 (2001) 183–203.
- [32] N. Klein, N. Epley, Group discussion improves lie detection, *Proc Natl Acad Sci U S A* 112 (2015) 7460–7465.
- [33] C.H. Lee, S. Cook, J.S. Lee, et al., Comparison of two meta-analysis methods: inverse-Variance-weighted average and weighted sum of Z-scores, *Genomics Inform* 14 (2016) 173–180.
- [34] Y. Nam, G.E. Park, J. Kang, et al., Fully automatic assessment of background Parenchymal enhancement on breast MRI using machine-Learning models, *J Magn Reson Imaging* 53 (2021) 818–826.
- [35] K. Borkowski, C. Rossi, A. Ciritis, et al., Fully automatic classification of breast MRI background parenchymal enhancement using a transfer learning approach, *Medicine (baltimore)* 99 (2020) 1–7.
- [36] S. Eskreis-Winkler, E.J. Sutton, D. D'Alessio, et al., Breast MRI background Parenchymal enhancement categorization using deep Learning: outperforming the radiologist, *J Magn Reson Imaging* 56 (2022) 1068–1076.