Detection of potential microcalcification clusters using multivendor for-presentation digital mammograms for short-term breast cancer risk estimation

Maya Alsheh Ali,[a] Mikael Eriksson, Kamila Czene, Per Hall, and Keith Humphreys
Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm SE-17177, Sweden

(Received 22 July 2018; revised 25 January 2019; accepted for publication 30 January 2019; published 7 March 2019)

Purpose: We explore using the number of potential microcalcification clusters detected in for-presentation mammographic images (the images which are typically accessible to large epidemiological studies) a marker of short-term breast cancer risk.
Methods: We designed a three-step algorithm for detecting potential microcalcification clusters in for-presentation digital mammograms. We studied association with short-term breast cancer risk using a nested case control design, with a mammography screening cohort as a source population. In total, 373 incident breast cancer cases (diagnosed at least 3 months after a negative screen at study entry) and 1466 matched controls were included in our study. Conditional logistic regression Wald tests were used to test for association with the presence of microcalcifications at study entry. We compared results of these analyses to those obtained using a Computer-aided Diagnosis (CAD) software (VuComp) on corresponding for-processing images (images which are used clinically, but typically not saved).
Results: We found a moderate agreement between our measure of potential microcalcification clusters on for-presentation images and a CAD measure on for-processing images. Similar evidence of association with short-term breast cancer risk was found ($P = 1 \times 10^{-10}$ and $P = 9 \times 10^{-09}$, for our approach on for-presentation images and for the CAD measure on for-processing images, respectively) and interestingly both measures contributed independently to association with a short-term risk ($P = 9 \times 10^{-03}$ for the CAD measure, adjusted for our proposed method and $P = 1 \times 10^{-04}$ for our proposed method, adjusted for the CAD measure).
Conclusion: Meaningful measurement of potential microcalcifications, in the context of short-term breast cancer risk assessment, is feasible for for-presentation images across a range of vendors. Our algorithm for for-presentation images performs similarly to a CAD algorithm on for-processing images, hence our algorithm can be a useful tool for research on microcalcifications and their role on breast cancer risk, based on large-scale epidemiological studies with access to for-presentation images. *© 2019 The Authors. Medical Physics published by Wiley Periodicals, Inc. on behalf of American Association of Physicists in Medicine.* [https://doi.org/10.1002/mp.13450]

Key words: breast cancer risk, breast cancer risk, for-presentation format, Multivendor full-field digital mammography

1. INTRODUCTION

Mammography is the most used imaging modality for breast cancer screening. Mammograms are these days acquired with full-field digital mammography systems and are provided in both for-processing (the raw imaging data) and for-presentation (a postprocessed version of the raw data) image formats. For-presentation images are intended for visual assessment by the radiologists. To save storage space, typically only for-presentation images are stored. While mammographic imaging is primarily used for diagnostic purposes, it has other, important uses. Recent years have seen intensive efforts searching for relevant information from mammograms to assist the prediction of long-term breast cancer risk. Many factors that increase the long-term risk of breast cancer are now known, such as a family history of breast cancer, a late age of first birth, an early onset of first menstruation, a late age at menopause. A large number of genetic markers of susceptibility to the disease, both rare and common have also been identified.[1] Mammographic breast density, which represents the amount of fibroglandular tissue in the breast, is the only strongly established image-based long-term risk factor for breast cancer.[2] While factors such as those listed above may be useful for long-term individualized screening,[3] there is also a need to find factors associated with short-term breast cancer risk, for which there would be an immediate clinical relevance (recalling women to screening earlier if they have a high short-term risk). Image features may be especially important in this context.

Incorporating mammographic image-based markers into short-term breast cancer risk models is an evolving and exciting research field.[4] One approach to finding such image characteristics is to carry out studies which include both women diagnosed with breast cancer, and women who are healthy, and to compare the two groups in terms of their previous negative screens. This is what we do in the current article. In the context of short-term risk assessment there is no relevant gold-standard that can be derived solely from prediagnostic,

negative images (events subsequent to the images being taken are the focus of interest). In this context, the gold standard is instead whether the women will/will not be diagnosed with breast cancer in the near future. The underlying logic is that even if no detectable tumour is currently present, there may be characteristics in a woman's mammographic image that are indicative of an increased risk of breast cancer in the near future (the coming few years).

Studies have shown that false positive recall is associated with short-term risk of breast cancer[5,6] and that a computer-aided diagnosis (CAD)-generated false-positive detection can be used as a quantitative imaging marker for short-term breast cancer risk.[7] CAD algorithms typically include the assessment of the presence of suspicious microcalcifications. These algorithms have typically been developed by training against a radiologist's assessments of the presence of malignant microcalcifications. It has been shown that the presence of suspicious clustered microcalcifications on a negative screening is important for short-term risk prediction.[8] However, It has been suggested that even nonsuspicious microcalcifications are associated with breast cancer.[9,10] To understand more about the role of microcalcifications in short-term risk of breast cancer, large-scale epidemiology studies need to be carried out and these will likely require the availability of algorithms for detecting microcalcifications on for-presentation mammographic images.

In recent decades, much research has been carried out to develop computer-aided algorithms, either for the detection of potential microcalcification (CADe) or for the analysis and diagnosis of microcalcification (CADx) and such algorithms are continually being developed. Algorithms for detection have primarily been developed for film-scanned (analog) images[11,12] or for raw (for-processing) digital images.[13,14] Different methods have been investigated, but, roughly speaking, these can be classified as either enhancement-based methods, multiscale analyses, classifier-based methods, or as mixtures of these methods,[15]. The majority of these methods consist of several stages: initial stages are focused on preprocessing and finding individual microcalcification candidates, and subsequent stages are aimed at grouping candidates into clusters and at removing false-positive clusters.[16] Although CAD has been used extensively in the United States to aid radiologists, within Europe CAD it is not implemented in screening programmes (because of its relatively low specificity), but instead, in most European countries, double reading is standard practice.

For a CADe system to be applicable to breast cancer screening it is important that its measurements are compatible even when mammograms are acquired with systems developed by different vendors. Each vendor has its own detector type for creating mammographic images, which results in substantial variations in noise characteristics and appearance. The automatic exposure control can be set to give a higher or lower dose for a particular breast thickness. Also each manufacturer's system uses a different image processing algorithm.

In this paper, we present a method that detects potential microcalcification clusters using for-presentation images from different digital systems. If an algorithm is robust across different digital systems, larger volumes of clinical data can be included in research projects. We evaluate our algorithm in an epidemiological study of the short-term risk of breast cancer using a matched case–control study design. In the present study, we have access to both for-presentation and for-processing images from the same examination. We are, therefore, able to compare the performance of our proposed method (on for-presentation images) to the performance of a CAD software[17] (on corresponding for-processing images). To the best of our knowledge, a study of this type, with an emphasis on microcalcification clusters in for-presentation images, has not before been presented in the literature.

2. METHODS

Microcalcifications are small calcium deposits that appear as white specks on a mammogram (with diameters between 0.1 and 1 mm, and average diameter of 0.3 mm[5]). The detection and segmentation of grouped microcalcifications is essentially a task of detecting arbitrarily shaped spots of a limited range of sizes. The main difficulties for reliable detection are the strong variations of the background that can introduce similar structures to microcalcifications, and the varying noise levels which are often hardly exceeded by signals; see Fig. 1. In addition, the linear structures in the mammogram, which can be ducts, blood vessels, or Cooper ligaments, often produce a characteristically textured appearance that may cause false detection. Intersection points sometimes appear as being more similar to calcifications than noise. Their identification and removal may improve the detection step. Their identification, however, remains a major challenge due to their large variety of widths, contrasts, lengths, and intersections.[18] Our proposed approach for detecting potential microcalcification clusters in for-presentation images can be divided into the following steps: (a) image preprocessing, consisting mainly of denoizing, quality improvement, and enhancement of small objects, (b) identification of potential microcalcifications; (c) filtering out potential true microcalcifications; and (d) grouping potential microcalcifications into clusters; see Fig. 2. We note that our algorithm (which is explained in detail in 2.C., below) was developed using 60 selected images from a large mammography cohort (for each of the different manufactures, we selected 20 images randomly; the image materials are described in Section 2.A, below). These images, used for developing our algorithm, were not from any of the individuals included in our case–control study/association analysis (statistical methods used to study association between case–control status and microcalcification clusters are described in Section 4, below).

2.A. Materials

Our study is based on digital mammograms from the Karolinska Mammography cohort (KARMA) study (`http://karmastudy.org/`), which is a prospective screening cohort that was initiated in January 2011 and comprises women attending mammography screening or clinical mammography at four hospitals in Sweden.[19] Participants

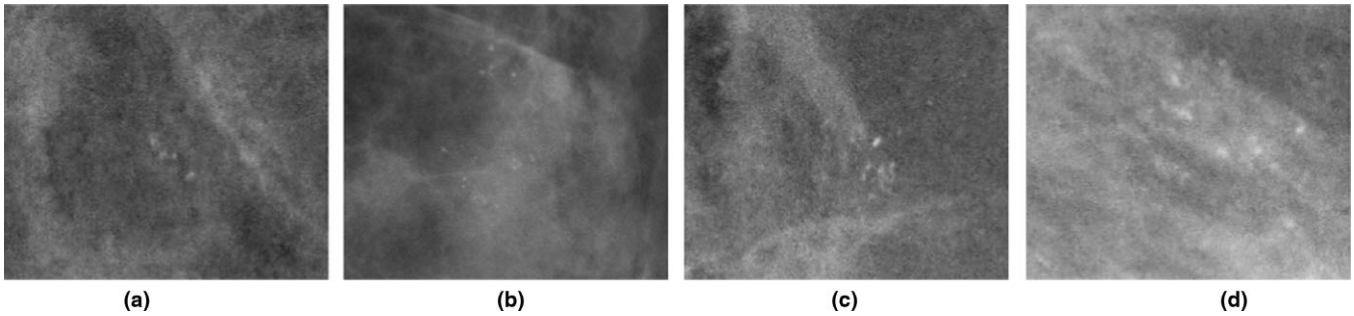

FIG. 1. Microcalcification clusters with different background: (a) fatty, (b) and (c) mixed with high density structure, and (d) glandular.

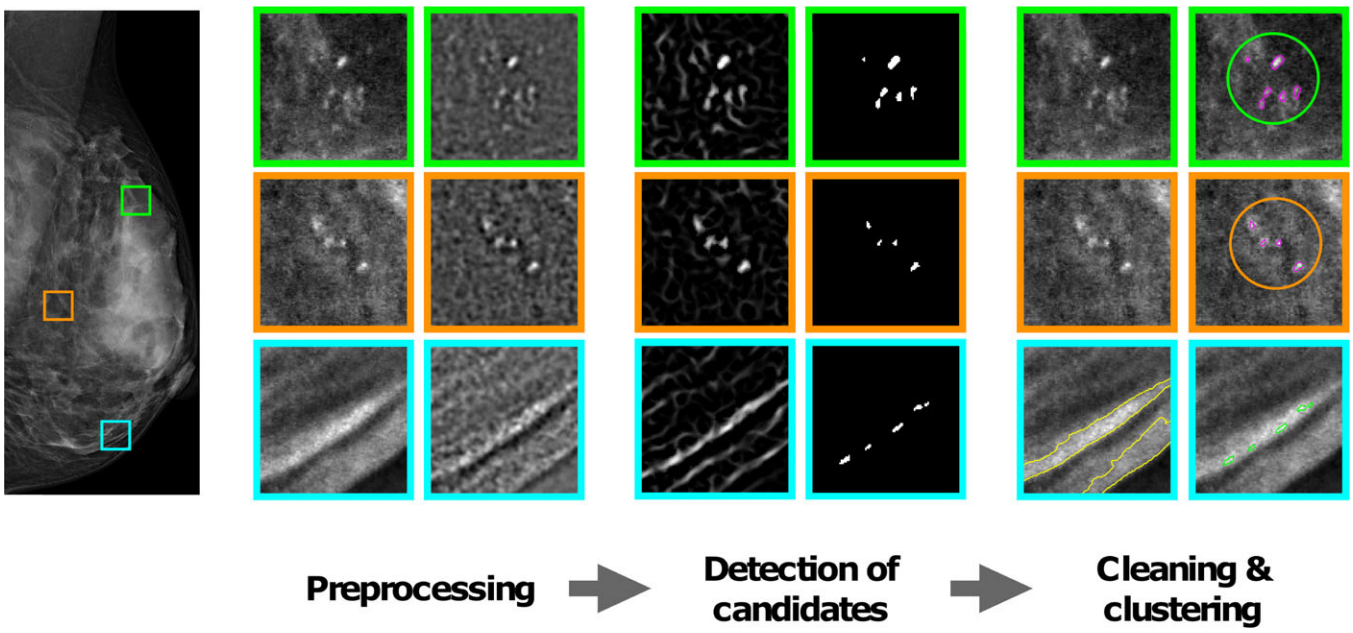

FIG. 2. A schematic overview of the proposed method. The preprocessing step includes the intensity transformation and the DOG filtering. The detection of candidates includes the HOG filtering and the segmentation. The final step includes removing noise and clustering of potential microcalcifications. The output of each step is shown on three different patches: a group of microcalcifications in the first two rows, and a false detection caused by linear structures in the third row. The microcalcifications in magenta are the retained ones and the ones in green are discarded in the cleaning step. Finally only the grouped microcalcifications are kept and the single microcalcifications are eliminated. [Color figure can be viewed at wileyonlinelibrary.com]

answered a comprehensive web-based questionnaire, allowed storage of mammograms, and accepted linkage to national breast cancer registers. By October 2015, a total of 570 incident breast cancers had been identified. From these cases we omitted women diagnosed with breast cancer within 3 months of a negative entry mammogram because it could not be excluded that their cancers were detected at the screening visit. Moreover, we required that full-field digital mammograms, both for-presentation and for-processing images, from the mediolateral oblique (MLO) view of the left and right breasts, were available. This left us with a total of 373 breast cancer cases (of which 331 are invasive cancer and 42 are in situ cancer). The median time between the used negative mammograms and the diagnosis date was 1.7 yr. We used the MLO view since it offers the best opportunity to visualize the maximum amount of breast tissue in a single image. Control subjects were individually matched to cases on machine vendor, age at mammography, and the number of screens in a prospective nested case–control design. All images were taken from the screen at study entry. We attempted to find four individually matched controls for each case, but in some instances we were only able to identify two or three (349 cases were matched with 4 controls each, 22 cases were matched with 3 controls each, and 2 cases were matched with 2 controls each). Data on area-based percent density (PD) (measured using the M-Vu CAD software), age, body mass index (BMI), hormone replacement therapy (HRT) status, parity, and age at first birth (AFB) were available for all included women. The value of PD for each woman was calculated as the average between the left and right breasts. The included cases overlapped to a large extent with those included in the study of Eriksson et al.[8], however, for the current study, we included new controls, as we wanted to match cases to controls with images from the same machine vendor (This was not necessary in Eriksson et al.[8] since, in that study, all image features were measured from

for-processing images). For the current study, we had images from three different manufacturers (722 GE medical system with pixel spacing of 0.1 mm, 609 Philips digital mammography with pixel spacing of 0.05 mm, and 508 Sectra Imtec AB with pixel spacing of 0.05 mm). The for-presentation images from the different digital systems have different image qualities and are based on different processing algorithms.

2.B. CAD algorithm

The commercial software VuComp M-Vu v6.3 was used to extract the number of potential microcalcification clusters from for-processing images. Since the techniques used by commercial vendors are proprietary, it is not possible to determine exactly which algorithm the system uses. A background is however provided in the US8855388B2 patent.[20] The algorithm uses an object detection process based on multiscale Laplacian of Gaussian filters where each scale is optimized to detect peaks in a certain absolute size range. The system uses probability density functions, in a three-stage classifier, to robustly model distributions for a variety of features at each stage, in order to classify microcalcifications. For the training phase, the patent reports that the object detection process and feature calculator were run on a training set containing a large number of radiographic images, with and without microcalcifications, with indications of type of calcification (e.g., malignant, benign, lucent, vascular, scratch (film handling artifacts), and tissue (not a calcification)) recorded by individuals with training in interpreting radiological images. As output, the CAD algorithm highlights only the potential suspicious microcalcification clusters.

2.C. Proposed method for detecting potential microcalcification clusters

2.C.1 Preprocessing

Although we are convinced that it is possible to detect potential microcalcifications with a reasonable degree of accuracy from for-presentation images, it needs to be recognized that these images do contain a considerable amount of high-frequency noise that can affect the detection of microcalcifications. Preprocessing steps of mammographic images usually consist of image denoizng, contrast enhancement, pectoral muscle removal, and breast segmentation from the background. Digital mammogram backgrounds are homogeneous which means that detection of the breast skin-line is a trivial task which can be carried out efficiently using simple thresholding methods. The microcalcifications are among the brightest objects in the image, so we wanted to spread out the upper gray levels (and contract the lower gray levels). To do this we applied intensity transformations using the cosine function. We let $I(x,y)$ denote the intensity of a two-dimensional image, then the transformed image is given by:

$$I'(x,y) = 1 - cos(\frac{\pi}{2} \times I(x,y)), \quad (1)$$

This process normalizes the intensity values to the range (0–1) at the same time.

The segmentation of the pectoral muscle is still a challenging issue in mammographic image analysis. Any error in this segmentation can lead to microcalcifications close to the pectoral muscle border being removed. This, together with the fact that microcalcificaitons can sometimes be found superimposed on the pectoral muscle, led us to choose not to remove the pectoral muscle prior to analysis.

Since the microcalcifications consist of groups of pixels, of a known size range, we can apply blob detection filters to enhance the microcalcifications, while removing both high-frequency components, representing image noise, and low-requency components that correspond to the homogeneous parts of the image. We applied the Difference of Gaussian filter (DoG), a band-pass filter in which both cut-off frequencies can be parametrized and are assumed to be associated to the potential microcalcification candidates and also to the noisy objects related to thin linear structures, such as ducts and Cooper ligaments. To limit the effects of noise and to maximize the enhancement of the microcalcifications, the standard deviations for both the low-pass and high-pass Gaussian kernels are set to be respectively the radius for the smallest possible size of a microcalcification (which is 0.05 mm) and the theoretical average size of a microcalcification (which is 0.15 mm).

2.C.2 Enhancement of linear structures and microcalcifications

The preprocessing step enhances average sized microcalcifications and removes most of the nonrelevant objects from the mammogram. However, the smaller microcalcifications have sizes and gray levels similar to both the noise and unwanted blobs from the linear structures. Therefore, some further processing is needed. We apply the Hessian of Gaussian (HoG) method to further enhance the brightness of structures with a specific shape, while encompassing all possible object sizes. By combining the eigenvalues of the HoG matrix at every pixel of the image, it is possible to define precise filters to extract both the microcalcifications of all biologically possible sizes and the linear structures.

The Hessian of $I(x,y)$ at pixel coordinates x and y with scale σ is represented by a 2×2 matrix:

$$H_\sigma(x,y) = \sigma^2 \times \begin{pmatrix} \frac{\delta^2}{\delta x^2} I(x,y) * G_\sigma(x,y) & \frac{\delta^2}{\delta x \delta y} I(x,y) * G_\sigma(x,y) \\ \frac{\delta^2}{\delta x \delta y} I(x,y) * G_\sigma(x,y) & \frac{\delta^2}{\delta y^2} I(x,y) * G_\sigma(x,y) \end{pmatrix} \quad (2)$$

where $G_\sigma(x,y) = (2\pi\sigma^2)^{-1} exp(-\frac{x^2+y^2}{2\sigma^2})$ is a bivariate Gaussian function and * denotes convolution. The information about the shapes of structures presented in the image and their brightness is reflected by the signs and magnitudes of the eigenvalues of the Hessian matrix H. Table I lists the possible combinations of shapes and brightnesses of the

TABLE I. Patterns in a 2D image that are identified from the analysis of Hessian eigenvalues, as a function of the values of the eigenvalues λ_k (H = high, L = low, N = noisy, ± indicate the sign of the eigenvalue).The eigenvalues are ordered: $|\lambda_1| \leq |\lambda_2|$.

λ_1	λ_2	Orientation pattern
N	N	Noisy, no preferred direction
L	H-	tubular structure (bright)
L	H+	tubular structure (dark)
H-	H-	blob-like structure (bright)
H+	H-	blob-like structure (dark)

structures that are identified in a 2D image by the analysis of Hessian eigenvalues ($|\lambda_{\sigma,1}| \leq |\lambda_{\sigma,2}|$).[21]

Both types of objects that need to be enhanced, blob-like and tubular structure, are bright. Therefore, all pixels with non-negative values of $\lambda_{\sigma,2}$ are removed before applying the enhancement function proposed by Jerman et al.[22]:

$$\varphi_\sigma = \lambda_{\sigma,2}^2 |\lambda_{\sigma,\rho}| \left(\frac{3}{2|\lambda_{\sigma,2}| + |\lambda_{\sigma,\rho}|} \right)^3, \quad (3)$$

where λ_ρ is a regularized value of λ_2 which ensures robustness of the enhancement function and is defined as:

$$\lambda_{\sigma,\rho} = \begin{cases} \lambda_{\sigma,2} & \text{if } \lambda_{\sigma,2} < \tau \times \min\{\lambda_{\sigma,2}\} \\ \tau \times \min\{\lambda_{\sigma,2}\} & \text{if } \tau \times \min\{\lambda_{\sigma,2}\} \leq \lambda_{\sigma,2} < 0 \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

where τ is a cutoff threshold between 0 and 1.

A multiscale filter response $F(x,y)$ is obtained by taking the mean of the given enhancement function v, at each point (x,y), over a range of scales σ, as

$$F(x,y) = \underset{\sigma_{\min} \leq \sigma \leq \sigma_{\max}}{\text{mean}} \{\varphi_\sigma\} \quad (5)$$

When the scale is small, unnecessary details can be highlighted, leading to over segmentation. When the scale is large, several small structures could be identified as one object, leading to under segmentation. Therefore, determining the optimum scale to extract most of the microcalcification clusters of interest is critical. In the present study, $\sigma_{\min}$ and $\sigma_{\max}$ were also set to 0.05 and 0.15 mm, respectively.

2.C.3 Segmentation of microcalcification candidates

The first part of the method, which is DoG, allows the approximate area of microcalcification to be determined, the contrast to be improved, and noise to be reduced in the image. Using HoG, we obtain a new image that emphasizes the approximate area of small microcalcifications. To segment the microcalcifications, an operating threshold $T = \mu + c\sigma$ is applied on both images, where μ and σ are, respectively, the mean and standard deviation determined from the corresponding image. Microcalcifications represent a tiny percentage of the mammogram surface. For this reason, we estimate that at most 0.5% of the mammogram pixels are sufficient to accommodate the entire population of microcalcification pixels. We consider also that the maximum number of microcalcifications that can be found in a mammogram is less than 200 microcalcifications. Thus, the parameter c is set to a value that fulfills the two previous criteria for each image. The final microcalcification candidates were obtained by the sum of the two thresholded images.

2.C.4 Postpruning and grouping

For filtering out noise, we use several criteria. The object contrast value, which is calculated on the original image (i.e., before preprocessing), is defined as the difference between the average gray level over the area of the object and the average over a 0.1-mm enlarged area around the same object. Microcalcifications are small objects that appear usually as circular or slightly elongated relatively bright spots in the mammogram. The objects with diameter less than 0.1 mm and more than 1 mm (which correspond to the smallest and largest size of a microcalcification) are discarded. The elongated objects, with eccentricity more than 0.95 are removed before calculating average gray levels of the two areas.

After the above filtering steps, there still, typically, exist a large number of disperse noise signals and agglomerate structural signals linked to the inhomogeneity of linear structures, bringing in substantial interference for the latter identification and analysis of microcalcifications. These structures can be removed from the detected microcalcification candidates since they lie along a linear structure/tubular object. To do this, we use the method described in Section 2.C.2, to generate an enhancement function with scales that correspond to the linear structures in the mammogram. We threshold the enhancement function using the operating threshold $T = \mu + c\sigma$, where μ and σ are the mean and standard deviation determined from the enhancement function. The value of the parameter c is set to a value such that only the most dense linear structures that cover 1% of the breast are retained. The obtained mask is then cleaned by removing all objects with a major axis length less than a certain threshold.

Isolated calcifications are not clinically significant, and thus a clustering criterion is usually incorporated into computerized detection methods. Thus, the last step in the cleaning process is microcalcification grouping. For this purpose, we use a Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm,[23] which groups together microcalcifications which are packed closely together, marking as outliers microcalcifications that lie alone in low-density regions. DBSCAN requires two parameters: the maximum distance between microcalcifications and the minimum number of microcalcifications required to form a dense region. Specifically, two microcalcifications are defined as belonging to the same group if their distance apart is less than 4.1 mm — and to form a group there should be at least four microcalcifications.

2.D. Statistical analysis

We evaluated the association (in terms of odds ratios and *P*-values) between the presence of microcalcification clusters and case–control status by fitting conditional logistic regression models treating case–control status as the dependent variable. First, we treated the total number of microcalcification clusters in the left and right breasts as an independent variable. Since in previous work in our group, Eriksson et al.[8] considered the absolute difference in the number of microcalcification clusters between the left and right breast as a predictor of short-term breast cancer risk, we also carried out analyses considering this as an independent variable. We included BMI, PD, HRT, parity, and AFB (the variables parity and AFB were combined into a single categorical variable with five categories) in all conditional logistic regression models as adjustment variables. We compared the results obtained by the proposed method with ones generated from the CAD algorithm. Because ours is a matched case–control study it is not straightforward to estimate a relevant measure of predictive modelling performance (pp.24,Pepe[24]). The standard "pooled" AUC applied to our data would describe the ability of the variables (BMI, PD, HRT, parity, AFB, and microcalcification variables) to distinguish between cases and age-matched controls, which would not be of practical interest (as it would not be population based) and would be attenuated because it would not include age. We instead, therefore, calculated age-adjusted AUCs using the covariate-adjusted ROC curve approach described in Pepe, Fan, and Seymour.[25] Honest values were calculated using the bootstrap procedure described by Harrell et al. We also compared the two outputs from the two methods measuring potential microcalcifications directly, by calculating a Kappa coefficient. We used Cohen's Kappa with squared weights,[26] which is appropriate for ordered categories (in this case numbers of microcalcification clusters).

3. RESULTS

We illustrate our algorithm for noise microcalcification detection, for an example image (from a Sectra digital mammography machine), in Fig. 2. The output of each step is shown on three different patches of the image. The background in this image is very nonhomogeneous. The first two patches both contain a group of fairly small calcifications, but with different contrasts. As a result, in the first patch some of the calcifications are missing. The third patch represents false detections caused by the nonhomogeneity of the linear structures and the way in which they were removed, using the mask of the linear structure in the cleaning step.

Key characteristics of cases and controls included in our association analyses are described in Table II. PD, calculated as the average between the left and right breasts, is, on average, higher in cases than in controls. This is not surprising since PD is known to be positively associated with breast cancer risk.[2] Summaries of the number of potential microcalcification clusters, in terms of both the total number of noise microcalcification clusters and the differences in the number of clusters between the left and right breasts, obtained by the proposed method, as well as by CAD using the raw images, are presented in Tables III and IV, respectively. In general, our approach finds more potential microcalcification clusters (in the for-presentation images) than CAD finds (in the for-processing images). There is, however, a moderate level of agreement between the two approaches (Cohen's Kappa using squared weights is 0.47); see Table V; and where CAD finds more than one cluster, our approach finds at least one cluster more than 75% of the time.

After adjusting for potential confounding variables, there is strong evidence of an association between the CAD measurement of the total number of potential microcalcification clusters and short-term breast cancer risk ($P = 9 \times 10^{-09}$). A very similar level of association ($P = 1 \times 10^{-10}$) was observed when using our method to calculate the total number of potential microcalcification clusters for the for-presentation images. The per-cluster odds ratios are slightly smaller (and confidence intervals slightly narrower) for our proposed

TABLE II. Key characteristics of individuals included in the breast case–control study. Means with standard deviations are given for continuous variables and proportions are given for categorical variables. *P*-values are obtained using likelihood ratio tests based on fitting conditional logistic regression models without adjustment for additional covariates.

Characteristic	Cases	Controls	*P*-value
Number	373	1466	
HRT use			0.169
Never	290 (77.5%)	1147 (78.2%)	
Past	62 (16.6%)	263 (18%)	
Current	22 (5.9%)	56 (3.8%)	
Parity and AFB			0.100
Nulliparous	47 (12.6%)	193 (13%)	
Parity ≤2 and AFB ≤25	97 (26%)	363 (24.8%)	
Parity ≤2 and AFB >25	149 (40%)	523 (35.7%)	
Parity >2 and AFB ≤25	39 (10.4%)	227 (15.5%)	
Parity >2 and AFB >25	42 (11%)	160 (11%)	
Age at mammography	57.286 (±9.128)	57.280 (±9.112)	0.176
BMI	25.570 (±4.613)	25.349 (±4.184)	0.392
PD	25 (±19.767)	19.622 (±18.420)	9×10^{-08}

PD, percent density; BMI, body mass index; HRT, hormone replacement therapy; AFB, age at first birth.

TABLE III. Total numbers of detected potential microcalcification clusters (# clusters) according to the CAD measure and our proposed method. Numbers, with (row) percentages in parentheses.

# clusters	0	1	2	3+
CAD				
Cases	272 (73%)	70 (18.7%)	19 (5.1%)	12 (3.2%)
controls	1291 (88.1%)	121 (8.2%)	35 (2.4%)	19 (1.3%)
proposed				
Cases	208 (55.8%)	91 (24.4%)	37 (9.9%)	37 (9.9%)
controls	1097 (74.8%)	238 (16.2%)	64 (4.4%)	67 (4.6%)

TABLE IV. Absolute differences in number of potential microcalcification clusters between the left and right breasts (L-R difference) according to the CAD measure and our proposed method. Numbers, with (row) percentages in parentheses.

L-R difference	0	1	2	3+
CAD				
Cases	275 (73.8%)	75 (20%)	18 (4.8%)	5 (1.4%)
controls	1304 (88.9%)	128 (8.7%)	26 (1.8%)	8 (0.6%)
proposed				
Cases	227 (60.6%)	101 (27.1%)	38 (8.3%)	15 (4.0%)
controls	1140 (77.7%)	262 (17.9%)	34 (2.6%)	26 (1.8%)

TABLE V. The confusion matrix for the total number of detected potential microcalcification clusters in both left and right breasts according to both CAD and the proposed method.

	Proposed			
# of clusters	0	1	2	3+
CAD				
0	1231	250	53	30
1	65	69	30	27
2	6	8	13	27
3+	4	2	5	20

method (more clusters are detected using our method). In analyses stratified on machine manufacturer, all associations were significant ($P < 0.05$) in all strata (both for the CAD measure and for our method); Table VI. We note that when including both the CAD measure and our measure of the total number of microcalcification clusters, in the same logistic regression model, the, *P*-values of association were 9×10^{-3} and 1×10^{-4}, respectively, indicating that each measure contributes important information, in addition to the other, in terms of short-term cancer risk. Using a bootstrapping procedure (based on 1000 bootstrap samples), we obtained honest estimates (Harrell et al.[27]) of 0.642 and 0.638 for the age-adjusted AUCs for the full model, using the total number of potential microcalcification clusters measured by our method and using CAD, respectively. (The corresponding apparent age-adjusted AUCs were 0.657 and 0.653). The honest estimate of age-adjusted AUC for the model excluding the number of microcalcification clusters was 0.607 (the apparent age-adjusted AUC was 0.624).

Using the absolute difference in numbers of clusters between the left and right breast gave similar results to using the total number of microcalcification clusters, as an independent variable. There was strong evidence of an association with short-term risk for both the CAD measure for the for-processing images ($P = 3 \times 10^{-10}$) and our measure for the for-presentation images ($P = 4 \times 10^{-10}$). Again, in analyses stratified on machine manufacturer, associations were significant ($P < 0.05$) in all strata (both for the CAD measure and for our method); Table VII. When including both our measure and the CAD measure of the difference in the number of microcalcification clusters, in the same logistic regression model, the, *P*-values of association were 1×10^{-05} and 2×10^{-05}, respectively. In the above analysis, differences in the number of microcalcification clusters and the total number of microcalcification clusters were treated as continuous variables.

We also carried out an analyses of association, for both measures of potential microcalcification clusters, based on the CAD measure and on our proposed approach, stratified on invasiveness of the tumour (invasive/in situ); see Table VIII. Not surprisingly, the odds ratios (point estimates) were larger for in situ cancer, but, importantly, it can be seen that measurements of potential microcalcification clusters are significantly associated ($P < 0.05$) with breast cancer when only invasive breast cancer is considered (and estimates of odds ratios do not differ much from those presented in Tables VI and VII, when all cases are considered).

4. DISCUSSION

We have developed a method for the detection of potential microcalcifications in for-presentation digital images and have shown, using multivendor mammograms, that its measure gives similar evidence of association with short-term risk of breast cancer to that obtained using an established CAD measure on for-processing images. Our study, therefore, demonstrates that algorithms identifying potential microcalcification clusters (which do not distinguish between suspicious and nonsuspicious subtypes) have the potential to be useful for short-term breast cancer risk assessment, and that

TABLE VI. Results of tests of association (odds ratios (OR) with 95% confidence intervals and *P*-values), overall and stratified on digital system, based on the proposed method and on CAD, between total number of potential microcalcification clusters in the left and right breasts and case–control status. Wald tests from logistic regression models with adjustment for potential confounders.

	CAD		Proposed method	
Machine	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
GE	1.813 (1.338,2.457)	1×10^{-04}	1.363 (1.083, 1.716)	8×10^{-03}
Philips	1.554 (1.111, 2.174)	1×10^{-02}	1.55 (1.248, 1.927)	7×10^{-05}
Sectra	1.763 (1.293, 2.404)	3×10^{-04}	1.774 (1.377, 2.286)	9×10^{-06}
all	1.685 (1.410, 2.014)	9×10^{-09}	1.529 (1.342, 1.742)	1×10^{-10}

TABLE VII. Results of association tests (odds ratios (OR) with 95% confidence intervals and *P*-values), overall and stratified on digital system, based on the proposed method and on the CAD measure, between differences in number of potential microcalcification clusters between left and right breasts and case–control status. Wald tests from logistic regression models with adjustment for potential confounders.

	CAD		Proposed method	
Machine	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
GE	2.178 (1.522, 3.118)	2×10^{-05}	1.701 (1.273, 2.273)	3×10^{-04}
Philips	1.860 (1.290, 2.683)	8×10^{-04}	1.483 (1.144, 1.923)	3×10^{-03}
Sectra	2.038 (1.353, 3.071)	6×10^{-04}	1.988 (1.438, 2.749)	9×10^{-06}
all	1.968 (1.593, 2.431)	3×10^{-10}	1.671 (1.421, 1.964)	4×10^{-10}

TABLE VIII. Association (odds ratios (OR) with 95% confidence intervals and p-values) between short-term risk of breast cancer and measures of the total number of potential microcalcification clusters and the absolute difference in numbers of clusters based on the CAD measure (for-processing images) and our proposed method (for- presentation images) stratified on invasiveness (invasive (n = 332) vs in situ (n = 42)).

	In situ		Invasive	
Method	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
total				
CAD	4.185 (1.862, 9.404)	5×10^{-04}	1.600 (1.326, 1.930)	2×10^{-07}
proposed	2.202 (1.296, 3.739)	3×10^{-03}	1.510 (1.317, 1.733)	2×10^{-09}
diff				
CAD	4.985 (1.907, 13.030)	1×10^{-03}	1.873 (1.499, 2.338)	3×10^{-08}
proposed	3.525 (1.778, 6.991)	3×10^{-04}	1.619 (1.366, 1.918)	2×10^{-08}

it is possible to develop algorithms which can be used on for-presentation mammograms.

Several studies have already investigated and compared the use of for-presentation and for-processing images for mammographic percent density estimation,[28,29] but little has been done for microcalcification detection. Recently, however, Wang et al.[30] investigated the feasibility of using for-presentation images in computerized analysis and diagnosis of microcalcification lesions. They found close agreement between for-presentation and for-processing images in terms of quantitative image features pertinent to microcalcification lesions and that a slight increase in false positives in MC detection is observed in for-presentation images. Their study was evaluated on a set of 188 matched mammogram pairs all containing clustered microcalcifications where image regions of interest centered around the micorcalcification clusters were cropped. All mammograms were acquired using a single vendor (GE). Our results demonstrate that it is possible to derive measures with a moderate degree of agreement between the two image formats, even when their focus is not the same. In our study, we detected, on average, more microcalcification clusters using our algorithm than the CAD measure (on for-processing images) that we used. This is likely to be primarily because of the difference in the preprocessing between the two image formats, and because we include all microcalcification candidates. There may, however, also be a higher level of false-positives in microcalcification detection based on for-presentation images, which would be in line with the results obtained by Wang et al.[31] In addition, the sensitivity of each method depends on the choice of parameters, such as the grouping criteria of microcalcifications, which can affect the obtained number of clusters.

Automated CAD systems (for for-processing images) are often designed with two stages: (a) detection of microcalcification candidates in the mammogram and (b) classifying each microcalcification group as benign or malignant (this is done by training against Radiologists' assessments). CAD systems perform these two steps and at the end keep only suspicious microcalcification clusters. In this study, we focused on the detection of potential microcalcification candidates using the for-presentation images from multi vendors. It is, however, possible that even potential nonsuspicious (benign) microcalcifications are useful for risk prediction. After all, we did obtain similar evidence of association with short-term risk using our approach as when using the CAD measure. Developing such algorithms is important because large epidemiological studies for deriving markers of (clinical) short-term risk prediction of breast cancer may have to be based on banks of for-presentation images.

There are, of course, limitations to any detection process, since it is not possible to (fully) account for random artifacts and noise that occur during the acquisition step. In addition, the detection of microcalcification clusters depends on the preprocessing of the raw image format and the used acquisition machine. Several studies in the literature have shown that the image processing step has a significant impact on the detection of microcalcifications in digital mammograms,[32] and that microcalcifications detection is sensitive to image quality and the dose used.[33]

The main challenge in working with for-presentation mammographic images of different digital systems is that there is no standardization. There is a need to develop post-processing methods for correcting for systematic variation in intensities due to the use of different manufacturer algorithms and their different versions through transforming all images from the given image gray scale into a standard gray scale wherein similar intensities achieve similar tissue meanings. This would make it easier to develop a robust algorithm.

Future work (with for-presentation images) investigating the association of features extracted from microcalcification clusters, such as the location of the cluster in the breast or its morphology, with case–control status would be of great interest.

5. CONCLUSION

We have presented a method for potential microcalcification detection using for-presentation images. The segmentation of potential microcalcifications in mammographic images is challenging due to their physical properties and to the noisy background of the image. We tackled the problem by first applying a selective detector, targeting both blob-like and line-like structures, to enhance the brightness of microcalcifications and to simultaneously obtain more homogeneous linear structures. The detection of linear objects is necessary to avoid detection errors linked to nodes in the vascular network or other noisy blobs of pixels. All the candidate

objects with shapes, sizes, and appearances similar to microcalcifications were then detected and further filtered to reduce the noise.

Our results demonstrate that it is possible to use for-presentation images for generating an image marker based on potential microcalcification cluster detection , which can facilitate large-scale epidemiological and clinical studies focused on modeling and understanding the short-term/clinical risk of breast cancer.

ACKNOWLEDGMENTS

We thank all the participants in the KARMA study and the study personnel for their devoted work during data collection. We are also grateful to the anonymous reviewers, the associate editor and the editor for detailed comments and suggestions on earlier versions of the manuscript.

FUNDING

This work was supported by the Swedish Research Council (grant number 2016-01245), the Swedish Cancer Society (grant number CAN 2017/287), the Swedish e-Science Research Centre, and the Cancer Health Risk Prediction Centre (CRISP; `www.crispcenter.org`), a Linneus Centre (contractID 70867902) financed by the Swedish Research Council.

CONFLICT OF INTERESTS

The authors have no conflicts to disclose.

[a)]Author to whom correspondence should be addressed. Electronic mail: maya.alsheh.ali@ki.se.

REFERENCES

1. Michailidou K, Beesley J, Lindstrom S, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat Genet*. 2015;47:373.
2. Boyd NF, Guo H, Martin LJ, et al. Mammographic density and the risk and detection of breast cancer. *New Eng J Med*. 2007;356:227–236.
3. Hall P, Easton D. Breast cancer screening: time to target women at risk. *Br J Cancer*. 2013;108:2202–2204.
4. Tan M, Zheng B, Leader JK, Gur D. Association between changes in mammographic image features and risk for near-term breast cancer development. *IEEE Trans Med Imaging*. 2016;35:1719–1728.
5. McCann J. Stockton D. Godward S. Impact of false-positive mammography on subsequent screening attendance and risk of cancer. *Breast Cancer Res*. 2002;4:R11.
6. Castells X, Roman M, Romero A, et al. Breast cancer detection risk in screening mammography after a false-positive result. *Cancer Epidemiol*. 2013;37:85–90.
7. Mirniaharikandehei S, Zarafshani A, Heidari M, Wang Y, Aghaei F, Zheng B. Applying a cad-generated imaging marker to assess short-term breast cancer risk. In: *Medical Imaging 2018: Computer-Aided Diagnosis*. Vol. 10575. Bellingham, WA: International Society for Optics and Photonics; 2018:105753F.
8. Eriksson M, Czene K, Pawitan Y, Leifland K, Darabi H, Hall, P. A clinical model for identifying the short-term risk of breast cancer. *Breast Cancer Res*. 2017;19:29.
9. Naseem M, Murray J, Hilton JF, et al. Mammographic microcalcifications and breast cancer tumorigenesis: a radiologic-pathologic analysis. *BMC cancer* 2015;15:307.
10. Wilkinson L, Thomas V, Sharma N. Microcalcification on mammography: approaches to interpretation and biopsy. *Br J Radiol*. 2016;90:20160594.
11. Ciecholewski M. Microcalcification segmentation from mammograms: a morphological approach. *J Digit Imaging*. 2017;30:172–184.
12. Jing H, Yang Y, Nishikawa RM. Detection of clustered microcalcifications using spatial point process modeling. *Phys Med Biol*. 2010;56:1.
13. Bria A, Karssemeijer N, Tortorella F. Learning from unbalanced data: a cascade-based approach for detecting clustered microcalcifications. *Med Image Anal*. 2014;18:241–252.
14. Mordang JJ, Janssen T, Bria A, Kooi T, Gubern-Mérida A, Karssemeijer N. Automatic microcalcification detection in multi-vendor mammography using convolutional neural networks. In: *International Workshop on Digital Mammography*, Springer, 2016:35–42.
15. Elter M, Horsch, A. Cadx of mammographic masses and clustered microcalcifications: a review. *Med Phys*. 2009;36:2052–2068.
16. Giger ML, Karssemeijer N, Schnabel JA. Breast image analysis for risk assessment, detection, diagnosis, and treatment of cancer. *Annu Rev Biomed Eng*. 2013;15:327–357.
17. M-Vu CAD, CAD, Nashua,NH, USA.http://icadmed.com.
18. Linguraru MG, Marias K, English R, Brady M. A biologically inspired algorithm for microcalcification cluster detection. *Med Image Anal*. 2006;10:850–862.
19. Gabrielson M, Eriksson M, Hammarstrm M, Borgquist S, Leifland, K, Czene, K, Hall, P. Cohort profile: the karolinska mammography project for risk prediction of breast cancer (karma). *Int J Epidemiol*. 2017;46:1740–1741.
20. Jeffrey CW, James PM, David SH, James HP, Anbinh, TH, Lawrence, MH. United States Patent No. US8855388B2. Retrieved from https://patents.google.com/patent/US8855388B2/en
21. Frangi AF, Niessen WJ, Vincken KL, Viergever MA. Multiscale vessel enhancement filtering. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 1998:130–137.
22. Jerman T, Pernuš F, Likar B, Špiclin, Ž. Enhancement of vascular structures in 3d and 2d angiographic images. *IEEE Trans Med Imaging*. 2016;35:2107–2118.
23. Ester M, Kriegel H-P, Sander J, Xu, X, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Kdd*, vol. 96, 1996:226–231.
24. Pepe, MS. *The statistical evaluation of medical tests for classification and prediction*. Oxford, UK: Oxford University Press; 2003.
25. Pepe MS, Fan J, Seymour CW. Estimating the receiver operating characteristic curve in studies that match controls to cases on covariates. *Acad Radiol*. 2013;20:863–873.
26. Fleiss JL, Cohen J, Everitt B. Large sample standard errors of kappa and weighted kappa. *Psychol Bull*. 1969;72:323.
27. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15:361–387.
28. Cheddad A, Czene K, Eriksson M, Li J, Easton D, Hall, P, Humphreys K. Area and volumetric density estimation in processed full-field digital mammograms for risk assessment of breast cancer. *PLOS ONE*. 2014;9:1–10.
29. Keller BM, Nathan DL, Wang Y, et al. Estimation of breast percent density in raw and processed full field digital mammography images via adaptive fuzzy c-means clustering and support vector machine segmentation. *Med Phys*. 2012;39:4903–4917.
30. Wang J., Nishikawa RM, Yang, Y. Quantitative comparison of clustered microcalcifications in for-presentation and for-processing mammograms in full-field digital mammography. *Med Phys*. 2017;44:3726–3738.
31. Wang J. Yang Y. Nishikawa RM. Quantitative study of image features of clustered microcalcifications in for-presentation mammograms. In *Image Processing (ICIP)*, 2016 IEEE International Conference on, IEEE, 2016:3404–3408.
32. Zanca F, Jacobs J, Van Ongeval C, et al. Evaluation of clinical image processing algorithms used in digital mammography. *Med Phys*. 2009;36:765–775.
33. Warren LM, Mackenzie A, Cooke J, et al. Effect of image quality on calcification detection in digital mammography. *Med Phys*. 2012;39:3202–3213.