



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/151863/>

Version: Published Version

Article:

Evans, Karla, Kulpan, Anne-Marie and Wolfe, Jeremy (2019) Detecting the “gist” of breast cancer in mammograms three years before localized signs of cancer are visible. The British Journal of Radiology. 20190136. ISSN: 0007-1285

<https://doi.org/10.1259/bjr.20190136>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Received:
06 February 2019

Revised:
10 May 2019

Accepted:
14 May 2019

<https://doi.org/10.1259/bjr.20190136>

Cite this article as:

Evans KK, Culpán A-M, Wolfe JM. Detecting the “gist” of breast cancer in mammograms three years before localized signs of cancer are visible. *Br J Radiol* 2019; **92**: 20190136.

FULL PAPER

Detecting the “gist” of breast cancer in mammograms three years before localized signs of cancer are visible

¹KARLA K. EVANS, PhD, ²ANNE-MARIE CULPAN, PhD and ³JEREMY M. WOLFE, PhD

¹Psychology Department, University of York, York, United Kingdom

²Health Education England, Halifax, United Kingdom

³Harvard Medical School and Brigham and Women's Hospital, Boston, MA, USA

Address correspondence to: Karla K. Evans
E-mail: karla.evans@york.ac.uk

Karla K. Evans and Jeremy M. Wolfe have contributed equally to this study and should be considered as senior authors.

Objectives: After a 500 ms presentation, experts can distinguish abnormal mammograms at above chance levels even when only the breast contralateral to the lesion is shown. Here, we show that this signal of abnormality is detectable 3 years before localized signs of cancer become visible.

Methods: In 4 prospective studies, 59 expert observers from 3 groups viewed 116–200 bilateral mammograms for 500 ms each. Half of the images were prior exams acquired 3 years prior to onset of visible, actionable cancer and half were normal. Exp. 1D included cases having visible abnormalities. Observers rated likelihood of abnormality on a 0–100 scale and categorized breast density. Performance was measured using receiver operating characteristic analysis.

Results: In all three groups, observers could detect abnormal images at above chance levels 3 years prior

to visible signs of breast cancer ($p < 0.001$). The results were not due to specific salient cases nor to breast density. Performance was correlated with expertise quantified by the number of mammographic cases read within a year. In Exp. 1D, with cases having visible actionable pathology included, the full group of readers failed to reliably detect abnormal priors; with the exception of a subgroup of the six most experienced observers.

Conclusions: Imaging specialists can detect signals of abnormality in mammograms acquired years before lesions become visible. Detection may depend on expertise acquired by reading large numbers of cases.

Advances in knowledge: Global gist signal can serve as imaging risk factor with the potential to identify patients with elevated risk for developing cancer, resulting in improved early cancer diagnosis rates and improved prognosis for females with breast cancer.

INTRODUCTION

Breast cancer is the second leading cause of cancer deaths in females in the developed countries.¹ While screening mammography is the best available tool for early detection of cancer, sensitivity and specificity are lower than what is desirable,² with false negative rates of 20–30% and false positive rates of about 10% reported in North America.^{3,4} We seek to exploit perception of the “gist” of abnormality to improve performance.

The human visual system quickly extracts the global structure and statistical regularities from everyday scenes, allowing us to “get the gist” of our environment before selective attention captures the details.⁵ Anecdotal reports of experts, supported by eye-tracking and psychophysical measures, indicate that similar gist processing operations

occur in the assessment of a mammogram^{6,7} and, indeed, in other medical image perception tasks.⁸ Radiological images can be thought of as a specialized class of scenes and radiologists are medical experts who have learned to apply the processes of visual cognition to these unusual scenes.^{9,10} In a series of experiments, Evans and colleagues have demonstrated that expert radiologists can classify mammograms as normal or abnormal at above chance levels after just 500 ms exposure.¹¹ There may be two types of global processing. Kundel and Nodine propose that initial “global analysis” guides attention to lesions¹² and, under some circumstances, observers can localize lesions after a 500 ms exposure.¹³ However, in the Evans et al studies, experts separate abnormal from normal images at above chance levels without an ability to localize the lesion.¹¹ This non-localizable global gist signal represents a different type

of signal of abnormality. Perhaps the clearest evidence for the existence of a non-localizable global gist signal is that it can be detected in the breast contralateral to the lesion where, of course, there is nothing to localize.¹⁴ This signal is not correlated with breast density nor is it based on asymmetry between left and right breasts.¹⁴

Of course, radiologists would never screen mammograms using just this global gist signal. However, if this global gist signal could be detected prior to onset of a visible lesion, it could serve as an imaging risk factor whose detection could modulate subsequent management of a patient. Consequently, we ask whether the global gist signal is detectable years before the cancer presents as a localized actionable mammographic lesion. In previous work, we have reported evidence that this is possible. Brennan *et al*¹⁵ found that radiologists were able to detect gist of cancer in mammograms years before there are any overt signs of cancer when these make up one-fifth of the cases examined. In the present study, our aim was to replicate and extend those findings, testing the viability of this signal in different reading conditions. Specifically, here we test whether the ability to detect the gist signal differs across different expert populations given different training and screening practices in the USA and UK.

METHODS AND MATERIALS

Stimuli and apparatus

The stimuli consisted of 116 (Experiments 1A–1C) and 200 (Experiment 1D) bilateral, full-field digital mammograms. Mammograms of 1980 × 2294 pixels were downsized to 800 × 1000 pixels to fit the computer display. Mammograms, drawn from 70 patients from Bradford (UK) Teaching Hospitals NHS Foundation Trust, were anonymized, adhering to ethical research governance standards. The 35 patients whose prior exams were used as "abnormal" cases, had histologically verified visible and actionable cancer. At the time of diagnosis, visible abnormalities were "subtle" masses and architectural distortions as determined by the independent radiologists who acquired the cases. The "abnormal" prior images did not contain visible, localized signs of cancer. The 58 "abnormal" images (29 mediolateral oblique (MLO) views, 29 craniocaudal (CC)) shown to observers, were acquired 3 years prior to the mammograms that had revealed visible and actionable cancer (Table 1). Thus, these "abnormal" images would have been considered "normal" mammograms at the time, since, of course, no one would have known that these patients would later develop breast cancer. The 58 abnormal cases were intermixed with 58 normal mammograms (29 MLO, 29 CC), taken from patients who showed no sign of disease for at least 3 years after the images were acquired.

Experiment 1D included an additional 100 mammograms (50 normal and 50 abnormal with visible cancerous lesions) taken from 100 patients. The abnormal mammograms were a mixture of obvious and subtle masses, architectural distortions and calcifications. The sets of abnormal and normal images consisted of 25 MLO and 25 CC views. By mixing these cases of visible cancer with the priors from females who would later develop cancer, we could determine if the presence of visible disease on some cases would block detection of the gist of abnormality in the priors.

All the experiments were conducted on a Dell Precision™ M6500 laptop using MATLAB R2012b. The experiment was displayed on a 17" screen at a viewing distance of 53 cm. The display monitor had a resolution of 1440 × 900 (Dell, Round Rock, Texas.) and a refresh rate of 85 Hz. In clinical practice, images would be presented on a monitor of higher resolution, but the benefits of a clinical grade monitor are minimal in a 500 ms exposure.

Observers and procedure

The four experiments had institutional review board approval, and each was conducted with a different sample of observers. All participants had normal or corrected-to-normal vision and gave informed consent. The observers in Exp. 1A and 1D were recruited at the Radiological Society of North America annual meeting (USA). While for Exp. 1B and 1C, observers were recruited at NHS Trust Hospitals in north of England (UK). The sample sizes for the experiments were dictated by the availability of the observers (Table 2).

Study participants in Experiment 1A were 21 attending radiologists (10 female; average age 46 years) recruited during the RSNA 2016 meeting and all practicing in the USA. Study participants in Experiment 1B were 9 attending radiologists (8 female; average age 46 years), practicing and recruited in the UK. Study participants in Experiment 1C were 11 female reading radiographers (non-MD) specializing in breast imaging (average age 46 years) primarily engaged in active case reading in the UK National Health Service Breast Screening Program. Study participants in Experiment 1D were 18 attending radiologists (9 female; average age 49 years) recruited during the RSNA 2017 meeting and all practicing in the USA.

Exp.1A–1C differed only in the composition of the expert observer group. All observers viewed the same images. Half were mammograms acquired 3 years prior to the mammograms that had showed visibly actionable abnormalities. The other half were priors of normal cases. Order of images was randomized for each observer. After three practice trials, participants completed two blocks of 116 experimental trials in which they viewed bilateral mammograms. On each trial (Figure 1), a fixation cross appeared in the center of the screen for 500 msec followed by a 500 msec presentation of the images. After the brief presentation, observers saw a white outline of the previously presented breasts. Observers rated the likelihood of an abnormality on a scale from 0 (clearly normal) to 100 (clearly abnormal). In the second block of trials the observers gave a density rating on a 4-point scale after another 500 msec presentation of the same images in a different random order. The scale was modeled on the BIRADS density scale (1, fatty; 2, scattered fibroglandular; 3, heterogeneously dense; 4, extremely dense). Feedback was provided only for the three initial practice trials. We collected density scores in order to determine if abnormality scores were a proxy for density, a known risk factor for cancer. If our readers were going to base their abnormality score on an assessment of density, that assessment would have been based on their 500 msec exposure to the images.

Table 1. Specification about 58 abnormal cases whose prior mammograms acquired 3 years before any screen visible cancer was detected

Age at prior screening mammogram	Study reader had examinations for comparison when the prior was viewed	View of the prior presented	Lesion type when cancer detected 3 years later	Lesion size	Pathology	BIRAD Parenchymal density
65	YES	CC	MASS ILL DEFINED	35 mm	IDC, DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
64	YES	MLO	MASS, LOBULAR & SMOOTH	20 × 15 mm	IDC, DCIS	HETEROGENEOUSLY DENSE
64	NO	MLO	MASS ILL DEFINED	6 mm	DCIS	FATTY
63	1 YEAR EARLIER	MLO	MASS ILL DEFINED	35 mm	ILC	FATTY
70	3 YEARS EARLIER	MLO	MASS, IRREGULAR	14.5 mm	DCIS	FATTY
64	1 YEAR EARLIER	MLO	MASS OVAL & SMOOTH	8 mm	IDC, DCIS	FATTY
61	NO	MLO	MASS, IRREGULAR	11 mm	IDC, DCIS	FATTY
62	NO	MLO	ASYMMETRY	9 mm	IDC,	HETEROGENEOUSLY DENSE
70	YES	CC	MASS LOBULAR & IRREGULAR	9 mm	IDC, DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
62	NO	MLO	ASYMMETRY	10 mm	IDC	FATTY
66	NO	MLO	MASS IRREGULAR & DISTINCT	8.7 mm	DCIS WITH MICROINVASION	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
63	NO	CC	ASYMMETRY	7 mm	IDC	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
62	YES	MLO	FOCAL ASYMMETRY	4 mm	IDC, DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
58	NO	MLO	FOCAL ASYMMETRY	12 mm	IDC, DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
60	YES	CC	MASS IRREGULAR	17 mm	IDC, DCIS	FATTY
56	NO	MLO	MASS IRREGULAR	5 mm	DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
51	NO	MLO	FOCAL ASYMMETRY	17 mm	IDC, DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
57	YES	MLO	MASS OVAL & SPECULATED;	16 mm	IDC, DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
57	NO	MLO	MASS OVAL & INDISTINCT	6 mm	IDC, DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
65	YES	MLO	MASS IRREGULAR	20 mm	IDC, DCIS	EXTREEMLY DENSE
57	NO	MLO	MASS ROUND & IRREGULAR	7 mm	IDC, DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
47	NO	MLO	ASYMMETRY	4 mm	IDC, DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
74	NO	MLO	FOCAL ASYMMETRY	13 mm	INVASIVE WITH MIXED FEATURES	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
75	3 YEARS EARLIER	MLO	MASS IRREGULAR	17 mm	INVASIVE WITH MIXED FEATURES	HETEROGENEOUSLY DENSE
70	2 YEARS EARLIER	MLO	MASS ROUND & INDISTINCT	6 mm	IDC, DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
45	1 YEAR EARLIER	MLO	MASS ROUND & INDISTINCT	30 mm	IDC, DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY

(Continued)

Table 1 (Continued)

Age at prior screening mammogram	Study reader had examinations for comparison when the prior was viewed	View of the prior presented	Lesion type when cancer detected 3 years later	Lesion size	Pathology	BIRAD Parenchymal density
68	1 YEAR EARLIER	CC	ARCHITECTUAL DISTORTION	15 mm	IDC, DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
56	2 YEARS EARLIER	MLO	FOCAL ASYMMETRY	10 mm	DCIS WITH MICRO-INVASION	HETEROGENEOUSLY DENSE
77	2 YEARS EARLIER	MLO	MASS IRREGULAR	11 mm	DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
76	NO	MLO	MASS	17 mm	IDC, DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
64	1 YEAR EARLIER	MLO	MASS IRREGULAR	15 mm	INVASIVE WITH MIXED FEATURES, DCIS	HETEROGENEOUSLY DENSE
45	NO	MLO	ASYMMETRY	15 mm	DCIS WITH MICRO-INVASION	FATTY
66	1 YEAR EARLIER	MLO	MASS OVAL & IRREGULAR	27 mm	IDC, DCIS	HETEROGENEOUSLY DENSE
63	2 YEARS EARLIER	MLO	ASYMMETRY	16 mm	INVASIVE MIXED FEATURES, DCIS	HETEROGENEOUSLY DENSE
58	1 YEAR EARLIER	MLO	ASYMMETRY	13 mm	ILC	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
69	NO	CC	ARCHITECTUAL DISTORTION	6 mm	IDC	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
55	2 YEARS EARLIER	CC	CALCIFICATION	9 mm	DCIS	HETEROGENEOUSLY DENSE
52	1 & 2 YEARS EARLIER	CC	ARCHITECTUAL DISTORTION	20 mm	IDC, DCIS	FATTY
67	NO	CC	ARCHITECTUAL DISTORTION	13 mm	DCIS	HETEROGENEOUSLY DENSE
39	1 YEAR EARLIER	CC	ASYMMETRY	10 mm	INVASIVE MIXED FEATURES, DCIS	FATTY
67	1 & 2 YEARS EARLIER	CC	MASS IRREGULAR	15 mm	IDC, DCIS	FATTY
48	5 YEARS EARLIER	CC	MASS IRREGULAR	12 mm	DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
77	NO	CC	2 MASSES OVAL & IRREGULAR	23 mm	IDC, DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
77	1 & 2 YEARS EARLIER	CC	MASS OVAL & IRREGULAR	12 mm	IDC, DCIS	FATTY
43	2 YEARS EARLIER	CC	MASS IRREGULAR	48 mm	IDC, DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
68	2 YEARS EARLIER	CC	MASS IRREGULAR	4 mm	DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
57	1 YEAR EARLIER	CC	MASS OVAL	3 mm	DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
68	1 YEAR EARLIER	CC	FOCAL ASYMMETRY	6 mm	IDC, DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY

(Continued)

Table 1 (Continued)

Age at prior screening mammogram	Study reader had examinations for comparison when the prior was viewed	View of the prior presented	Lesion type when cancer detected 3 years later	Lesion size	Pathology	BIRAD Parenchymal density
56	1 & 2 YEARS EARLIER	CC	ASYMMETRY	6 mm	IDC	FATTY
64	1D YEAR EARLIER	CC	ASYMMETRY	15 mm	ILC AT TWO SIDES	FATTY
62	NO	CC	MASS	3 mm	DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
68	NO	CC	MASS ROUND & INDISTINCT	9 mm	IDC, DCIS	FATTY
57	YES	CC	ARCHITECTUAL DISTORTION	14 mm	DCIS	FATTY
68	YES	CC	MASS IRREGULAR	15 mm	IDC, DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
72	YES	CC	MASS ROUND & IRREGULAR	6 mm	IDC, DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
64	NO	CC	FOCAL ASYMMETRY	8 mm	IDC, DCIS	SCATTERED AREAS OF FIBROGLANDULAR DENSITY
56	NO	CC	ASYMMETRY	18 mm	ILC AT TWO SITES	SCATTERED AREAS OF FIBROGLANDULAR DENSITY

DCIS, ductal carcinoma in situ; IDC, invasive ductal carcinoma.

Experiment 1D mixed priors that would eventually develop cancer with cases with currently visible abnormalities. The procedure for Experiment 1D was otherwise similar to previous Experiments 1A–1C. 50 images of each type of abnormal case were intermixed in one block of 200 trials with 100 normal images.

STATISTICAL ANALYSIS

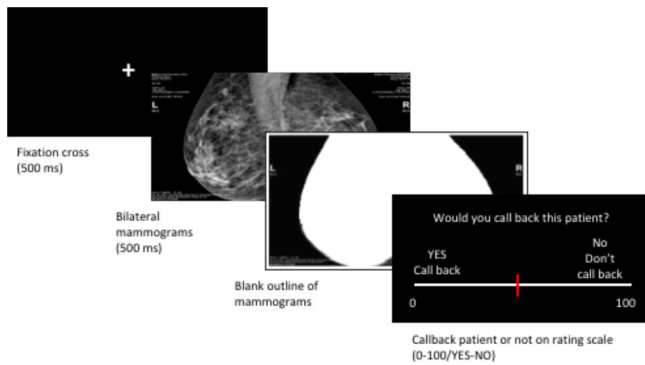
We converted the rating scale data to receiver operating characteristic (ROC) curves and calculated *d'* and area under the curve (AUC) measures. The statistical analysis was done on

the *d'* scores. ROCs can be calculated in two different ways, the conventional¹⁶ and using log linear likelihood ratios after smoothing (LLRs) to determine decision criterion.¹⁷ Because raw ratings tended to show bimodal distributions for normal and abnormal cases an optimal performance could not always be determined using a single criterion. Therefore, in addition to the conventional standard method we computed decision variables by first smoothing the raw data by fitting a Gaussian Kernel with bandwidth of 10 and calculating the log likelihood ratios to compute the AUC to characterize observer performance.¹⁷ In addition, the standard *d'* and AUC measures that we report

Table 2. Demographic data on observers who participated in Experiment

Observer group	Radiologist	Radiology residents	Reading radiographers	Years of experience	Percentage in breast imaging	Number of cases read in last year
Experiment 1A	17			22 (5–40)	44 (15–100)	4100 (1500–10,000)
		4		3 (1–4)	55 (10–100)	3200 (150–7000)
Experiment 1B	8			19 (9–33)	80 (50–100)	6200 (2000–9000)
		1		1	50	5000
Experiment 1C			11	18 (8–28)	100	5550 (1200–10,000)
Experiment 1D	17			20 (10–40)	60 (15–100)	3900 (700–8000)
		1		2	17	380

Figure 1. Experimental procedure for experiments 1 A-D.



assume equal variance for signal and noise distributions. This may not be a safe assumption for radiologic images.¹⁸ Accordingly, we also calculated $d(a)$ and A_z measures that do not rely on the equal variance assumption. The pattern of results does not change (Table 3). The item analysis of images used point-biserial correlations. Comparison between three expert groups' performance was done using an independent ANOVA. To examine the relationship between measures of expertise and performance we used simple linear regressions.

RESULTS

For 21 US (1A) and 9 UK (1B) radiologists, observers' ability to distinguish normal from abnormal (cancer priors) was modest in size but statistically significant (Exp.1A $d'=0.21$, s.e.m. = 0.05, $t(20) = 3.947$, $p = 0.0008$, AUC = 0.54, LLR AUC = 0.60; Exp.1B $d'=0.22$, s.e.m. = 0.06, $t(8) = 4.036$, $p = 0.0038$, AUC = 0.54, LLR AUC = 0.62), (Figure 2a,b). As can be seen, the LLR estimate of AUC's reported give somewhat larger values. The important point is that there is statistically significant evidence for the detectability of a global gist signal regardless of which method is used. An item analysis showed that that the results were not due to any specific, salient cases.

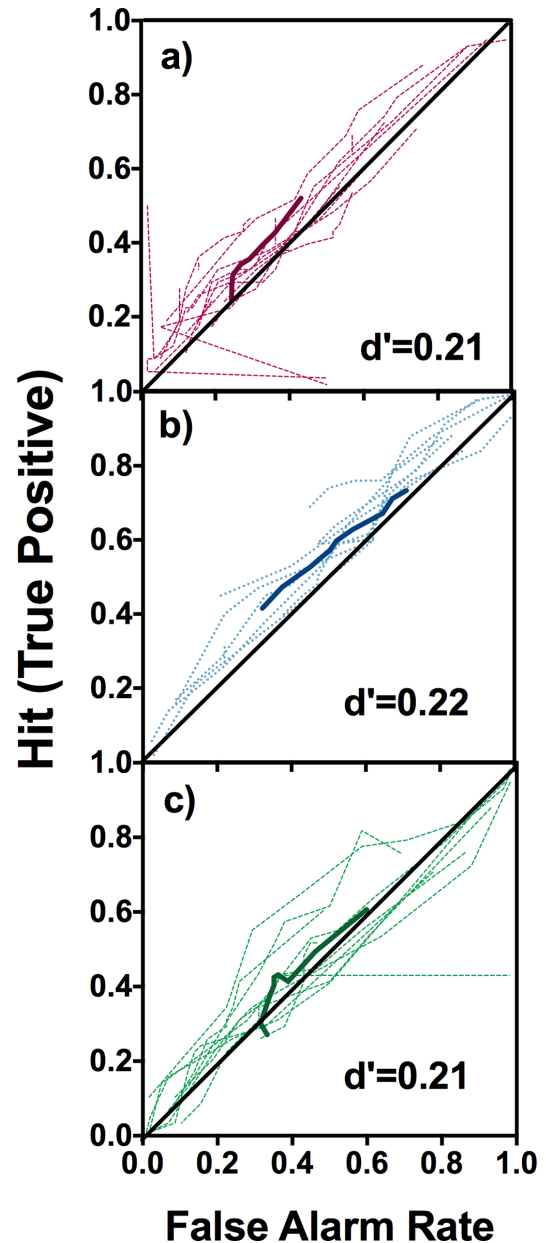
As noted above, we obtained density ratings of the 500 ms exposures [inter-rater reliability 1A intraclass correlation coefficient=0.645, 95% confidence interval (CI) (0.576 to 0.713) ($F(115,2300)=49.41$, $p<.001$); 1B intraclass correlation coefficient=0.558, 95% CI (0.485 to 0.635) ($F(115,920)=13.23$, $p<.001$)] in order to determine if the gist signal might be based on a rapid assessment of breast density. The data show that it is not. If this were the case, we would expect ratings of gist abnormality to increase with density. Instead, it is harder to detect the gist signal

Table 3. Average values for d' , AUC/LLC AUC, $d(a)$, and A_z for Experiments 1A-1C

	d'	AUC/LLC AUC	$d(a)$	A_z
US-Radiologist	0.21	0.54/0.60	0.78	0.71
UK-Radiologists	0.22	0.54/0.62	1.21	0.83
UK Radiographers	0.21	0.53/0.61	1.06	0.72

AUC, area under the curve.

Figure 2. ROC curves for the three observer groups of experiment 1A-C. Solid colored line, average ROC curve; light dotted lines, individual observers. (a) Performance of US radiologists at RSNA 2016 (b) Performance of UK radiologists (c) Performance of UK reading radiographers. ROC, receiver operating characteristic.



at high density. Using the data from Experiment 1A (US radiologists), we do find a correlation between rapid density ratings and abnormality ratings, but it is small (average correlation of 0.10; $t(20) = 3.51$, $p=0.0022$ two-tailed). Note that abnormality ratings run from 0-abnormal to 100-normal, so the correlation of 0.1 actually means that rated level of abnormality declines slightly as density increases. Looking at the data from Experiment 1B, we also find a significant correlation ($r=0.26$, $t(8) = 6.83$, $p=0.0001$ two-tailed). If we look at performance as a function of density rating, we find that d' increases modestly as a function of density

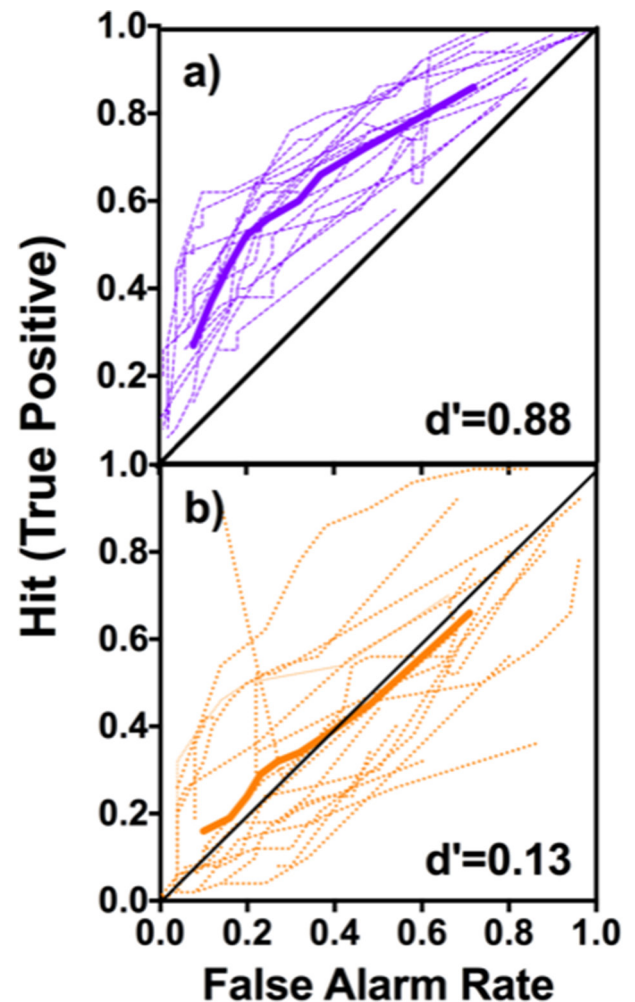
for moderate densities (aggregating data over all 21 observers and using a rating criterion of 50 to split the data; for density 1, $d' = 0.21$; density 2, $d' = 0.24$; density 3, $d' = 0.30$. At density=4, d' collapses to -0.28). Observers were unable to extract a gist signal from breasts in the highest density category. We repeated the d' analysis including data only if the observer rated that case as a density of 2. This eliminates about half of the data. With all cases having the same density rating, average $d' = 0.24$, $t(20) = 4.45$, $p < 0.0025$ (two tailed).

We also have standard density ratings for these images; those obtained without time restriction in the original clinical interpretation. The density ratings obtained in a flash correlate with those standard ratings (1A the Pearson $r = 0.40$, $t(20) = 34.89$, $p < .0001$; for 1B Pearson $r = 0.36$, $t(8) = 14.01$, $p < .0001$). In these experiments, observers would have access only to their impression of density in 500 ms. Still, it is interesting to note that the cases are rated as more likely to be *normal* as standard density increases, the opposite of what would be expected ($t(56) = 2.27$, $p = 0.027$) if the gist signal was a proxy for the standard density rating.

Exp. 1C was conducted with non-MD experts; radiographers who are trained to read mammograms and regularly participate in the breast screening program in the UK. Our aim was to determine if the ability to detect the global gist signal is due to primarily perceptual expertise that radiographers would have or whether it might depend on more the extensive medical knowledge possessed by radiologists. We compared the performance of the three expert groups and found no difference between their ability to detect mammograms of females that would go on to develop cancer 3 years later ($F(2,40) = .035$, $p = .966$). There was no difference between the UK and US radiologists (Gabriel's posthoc $p = .990$). More notably, the radiographers' performance was very similar to that of both the US (Gabriel's posthoc $p = .998$) and UK radiologists (Gabriel's posthoc $p = .875$), with a $d' = 0.21$, s.e.m. = 0.05, AUC = 0.53, LLR AUC = 0.61, significantly above chance ($t(10) = 4.253$, $p = .0017$, see Figure 2c).

If this gist signal were ever to be used in a clinical setting, it would be useful if it could be detected in prior exams of females who would develop cancer even when those priors were intermixed with cases of currently visible abnormality. Alternatively, it is possible that the stronger signals from visible abnormalities would effectively mask detection of weaker signals in the prior images. Thus, in Experiment 1D, mammograms collected 3 years prior to onset of cancer were intermixed with cases that had visible cancers (clearly visible, as well as subtle cases). Overall, the observers were well above chance at distinguishing cases with visible cancer from normal cases ($d' = 0.88$, s.e.m. = 0.08, $t(17) = 9.40$, $p < 0.0001$, AUC = 0.68, LLR AUC = 0.70; Figure 3a) replicating previous findings. However, unlike our findings in experiments 1A–1C, in this intermixed design, the observers in Experiment 1D were unable to reliably distinguish priors of cases that would develop cancer in 3–5 years from those that would remain normal for at least 3 years ($d' = 0.13$, s.e.m. = 0.17, $t(17) = 0.691$, $p = .499$, AUC = 0.48, LLR AUC = 0.49; Figure 3b). However, in a posthoc analysis, we looked at the performance of

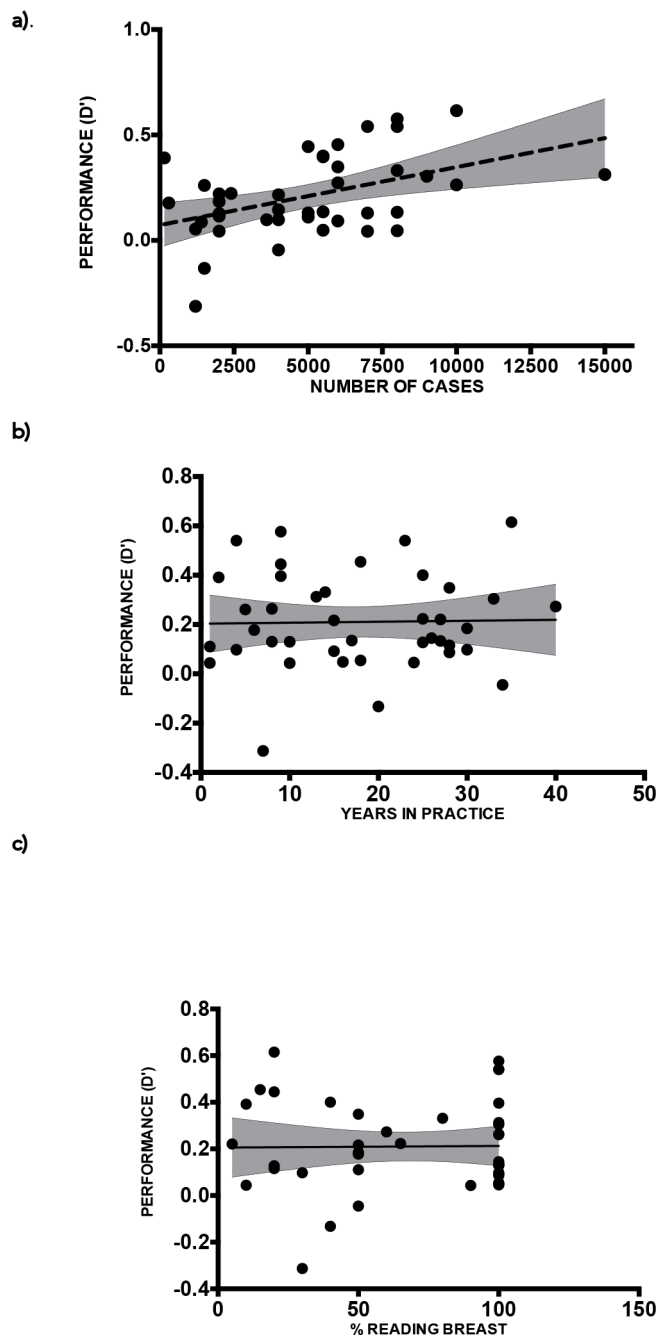
Figure 3. ROC curves for the US radiologists observer group in experiment 1D. Solid colored line, average ROC curve; light dotted lines, individual observers; dark dotted line. (a) Performance of observers in distinguishing cases of visible cancer from normal mammograms (b) Performance of observers when distinguishing of priors with no visible cancer but that would go on to develop cancer in 3 years from normal cases. ROC, receiver operating characteristic.



the six radiologists in this group, who devoted 100% of their time to breast imaging and who read 6000–8000 mammograms a year. These observers were able to distinguish priors of cancerous and normal cases surprisingly well ($d' = 1$, s.e.m. = 0.09, $t(5) = 10.96$, $p < 0.0001$, AUC = 0.69, LLR AUC = 0.72). This level of performance was similar to their performance with the cases of visible abnormality as visible cancers ($d' = 1.14$, s.e.m. = 0.09, $t(5) = 12.95$, $p < 0.0001$, AUC = 0.70, LLR AUC = 0.73). The two conditions were not significantly different in this group ($t(5) = 1.022$, $p = 0.354$). Since we separated this group of observers out after the fact, one would like to see this result replicated with a group of observers pre-selected for high expertise.

Clearly, the ability to see the "gist" of cancer is a learned skill (in earlier control experiments, novice observers performed at chance levels). In order to examine the effects of increasing

Figure 4. Observers' performance (d') across expert groups and experiments as a function of (a) the number of cases reviewed in the last year; (b) years of experience; (c) percentage of time reading mammograms.



experience, we examined the relationship of three measures of experience/expertise to performance on the gist task: (1) percentage of time devoted to breast imaging, (2) years of experience in imaging and (3) number of mammograms read each year. For this analysis, we combined data from all of the observer populations in the experiments described above. This seems justified, given that the different observer populations produced very similar performance. Thus, Figure 4 shows all 41 observers' performance (d') as a function of the number of cases reviewed in

the last year, years of experience, and percentage of time spent in breast imaging. The results showed that mammogram discrimination improved with number of cases reviewed, $F(1, 39) = 9.8$, $p = 0.0033$, $R^2 = 0.20$, 95% CI (-0.03 to 0.18), but not with years of experience, $F(1, 39) = .2$, $p = 0.8932$, $R^2 = 0.0004$, 95% CI (0.03, 0.32), nor percentage of time spent in breast imaging, $F(1, 39) = 0.006$, $p = 0.9376$, $R^2 = 0.0001$, 95% CI (0.07, 0.35).

DISCUSSION

The results, presented above, show that a global perceptual signal, related to the development of breast cancer, is visible at least 3 years before a local, actionable sign of cancer is present. Surprising as this may seem, this is plausible. The ability to extract semantic information from brief glimpses of scenes is well established.¹⁹ When the observer first sees a natural scene, its features are unbound²⁰⁻²² and its objects are not explicitly recognizable.²³⁻²⁵ Nevertheless, an observer can still extract quite a rich "gist" in a brief exposure.²⁶⁻²⁸ In a single glimpse (<200 ms, with a mask), observers can estimate average color, motion, size and orientation, for example.²⁹ They can categorize complex natural scenes (e.g. "beach," "office")^{19,30,31} and identify the presence of classes of objects (e.g. animal) though observers who correctly detect the gist of "animal," may not know the identity or location of that animal.^{19,25} With natural images, this ability appears to be based on classification of the raw feature statistics in the image.³² In mammograms, there is evidence for specific textural statistics associated with cancer.³³ Recent evidence suggests that the content of scenes is predominately conveyed by high spatial frequencies in the image.³⁴ Similarly, in previous studies, we have noted that the gist perceptual signal related to cancer is stronger in the high spatial frequencies.¹⁴ Given that many cancers may be associated with a genetic predisposition, it could well be that the genetics that predispose to cancer, also change the breast parenchyma in a manner that has perceptual consequences.

Further, we find that despite different screening/training practices in the USA and UK as well as across different expert reader populations we find no significant differences in signal sensitivity for the global gist signal in mammograms. It appears that the ability to detect this signal is driven primarily by perceptual expertise related to the number of images that have been seen.

One limitation of the current finding is that the measured gist signal is obviously quite small. However, the results shown here should be considered a conservative estimate of the potential of this signal. It is worth noting that these were 3 year prior images from a set of cancers that were deliberately chosen to be "subtle" at the time of diagnosis. Cases with calcifications or more obvious cancers were excluded since the original studies on the global gist signal did not use these types of cases. For detection of the gist of abnormality in prior images, the visibility of the cancer that eventually develops is not critical. In future work, it will be of interest to determine if the early-warning signal is larger for some types of breast cancer than for others. Different genetic subtypes do appear to generate different signals for computer vision algorithms. For example, a Bayesian Artificial Neural Network algorithm, can distinguish the appearance of the parenchyma in patients with or without BRCA1/2-related breast

cancer.^{35,36} Potentially, the gist signal that humans detect could be a marker for one or more genetic subtypes. This is a question that our current data cannot address but would be worthy of further investigation.

The signal might also be larger if observers could look at the image for a longer period of time. In our original gist studies, images were presented for a fraction of a second because it was important to minimize the possibility that the radiologist could search for and locate an actual lesion. With the prior images used here, there is nothing to search for. In future studies, prior images could be presented until the observer chooses to respond. In such a study, readers might be informed that 50% of the images came from females who would develop cancer within 3 years. Readers would then be asked to sort the images into normal and abnormal. Thus sensitized, readers might be able to find the gist of abnormality more successfully, given more time.

Another limitation to any application of this signal was observed in our Experiment 1D where a small gist signal in the priors seemed to be drowned out by stronger signals of visible cancer when both type of cases were read in a mixed batch, as would happen in real-life clinical practice. Unlike the previous report,¹⁵ we find that the signal is hard to find in the intermixed design. Experience may be the critical difference between these two studies. Our readers had a greater range of experience levels than the readers in the Brennan et al¹⁵ study, both in terms of percentage of time spent reading mammograms and the in number of mammograms read in a year. When we limited our analysis to the radiologists who read 6000–8000 cases/yr (approximating the expertise of the Brennan et al, readers), they were unimpaired; discriminating normal from abnormal priors just as effectively as they could discriminate normal from currently abnormal cases. The percentage of the time our experts devoted to breast imaging was correlated with the number of

mammogram cases they read ($r = 0.42$, $p = .006$) but did not predict their ability to detect the global signal. It was the annual number of cases read that appeared to be the basis for this expert behavior. This suggests that the ability to detect gist in priors could be learned through repeated exposure (in humans or machines). The role of number of cases read can be seen at the other end of the expertise scale, as well. The 15 readers who were the least reliable in our studies (as seen in Figures 2a & c, and 3b) all read less than 2500 cases a year.

The use of a stimulus set having a 50% cancer prevalence rate may limit the generalizability of our result. Prevalence is much lower in clinical screening practice. In our earlier work on the effects of low prevalence, we find that the primary effect is to make observers more conservative.³⁷ This might reduce the detection of gist abnormality at low prevalence though the effect of prevalence bias on gist detection remains to be studied. In any case, the gist signal is likely to remain fairly small; certainly, too small to be diagnostic in its own right. Gist seems most likely to be useful if treated as a risk factor, like breast density. No one would treat a patient based on breast density alone, but the risk factor of high density can change how a patient's screening is managed: higher risk triggering greater vigilance. The gist signal could be similarly useful in a personalized risk stratified care pathway for example. If it is proven to be useful, it is worth noting that gist is available with no additional screening or radiation exposure and with very little added demand on the clinician's time. The gist signal might also be a useful target for computer vision approaches. Deep learning methods are becoming increasingly common in radiology.^{38,39} Using such methods to detect a gist signal would be different from standard practice since the network would be trained to detect the warning sign, and not the actual visible disease. For the present, these results are evidence that there is a signal in some mammograms that is related to later development of cancer. Future work will reveal how useful this signal may be.

REFERENCES

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA: A Cancer Journal for Clinicians* 2013; **63**: 11–30. doi: <https://doi.org/10.3322/caac.21166>
2. Hubbard RA, Kerlikowske K, Flowers CI, Yankaskas BC, Zhu W, Miglioretti DL. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography. *Annals of Internal Medicine* 2011; **155**: 481–92. doi: <https://doi.org/10.7326/0003-4819-155-8-201110180-00004>
3. Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. *Radiology* 1992; **184**: 613–7. doi: <https://doi.org/10.1148/radiology.184.3.1509041>
4. Majid AS, de Paredes ES, Doherty RD, Sharma NR, Salvador X. Missed breast carcinoma: pitfalls and pearls. *RadioGraphics* 2003; **23**: 881–95. doi: <https://doi.org/10.1148/rg.234025083>
5. Wolfe JM, V6 MLH, Evans KK, Greene MR. Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences* 2011; **15**: 77–84. doi: <https://doi.org/10.1016/j.tics.2010.12.001>
6. Nodine CF, Kundel HL. The cognitive side of visual search in radiology. *Eye movements: From physiology to cognition* 1987; 573–82.
7. Kundel HL, Nodine CF, Conant EF, Weinstein SP. Holistic component of image perception in mammogram interpretation: Gaze-tracking study. *Radiology* 2007; **242**: 396–402. doi: <https://doi.org/10.1148/radiol.2422051997>
8. Krupinski EA, Tillack AA, Richter L, Henderson JT, Bhattacharyya AK, Scott KM, et al. Eye-movement study and human performance using telepathology virtual slides. Implications for medical education and differences with experience. *Human Pathology* 2006; **37**: 1543–56. doi: <https://doi.org/10.1016/j.humpath.2006.08.024>
9. Nodine CF, Mello-Thoms C. The nature of expertise in radiology. *Handbook of Medical Imaging*. SPIE 2000; 859–95.
10. Bertram R, Helle L, Kaakinen JK, Svedström E. The effect of expertise on eye movement behaviour in medical image perception. *PLoS ONE* 2013; **8**: e66169. doi: <https://doi.org/10.1371/journal.pone.0066169>
11. Evans KK, Georgian-Smith D, Tambouret R, Birdwell RL, Wolfe JM. The GIST of the abnormal: Above-chance medical decision making in the blink of an eye. *Psychon Bull*

- Rev 2013; **20**: 1170–5. doi: <https://doi.org/10.3758/s13423-013-0459-3>
12. Kundel HL, Nodine CF, Krupinski EA, Mello-Thoms C. Using gaze-tracking data and mixture distribution analysis to support a holistic model for the detection of cancers on mammograms. *Academic Radiology* 2008; **15**: 881–6. doi: <https://doi.org/10.1016/j.acra.2008.01.023>
 13. Carrigan AJ, Wardle SG, Rich AN. Finding cancer in mammograms: if you know it's there, do you know where? *Cognitive Research: Principles and Implications* 2018; **3**: 10. doi: <https://doi.org/10.1186/s41235-018-0096-5>
 14. Evans KK, Haygood TM, Cooper J, Culpan A-M, Wolfe JM. A half-second glimpse often LETS radiologists identify breast cancer cases even when viewing the mammogram of the opposite breast. *Proceedings of the National Academy of Sciences* 2016; **113**: 10292–7. doi: <https://doi.org/10.1073/pnas.1606187113>
 15. Brennan PC, Gandomkar Z, Ekpo EU, Tapia K, Trieu PD, Lewis SJ, et al. Radiologists can detect the 'gist' of breast cancer before any overt signs of cancer appear. *Scientific Reports* 2018; **8**: 8717. doi: <https://doi.org/10.1038/s41598-018-26100-5>
 16. Macmillan NA, Creelman CD. *Detection theory: A user's guide*: Psychology press; 2004.
 17. Semizer Y, Michel M, Evans KK, Wolfe J. *Texture as a Diagnostic Signal in Mammograms. CogSci 2018 Conference Proceedings*. Madison, WI, USA; July 25, 2018.
 18. Kundel HL. Disease prevalence and the index of detectability: a survey of studies of lung cancer detection by chest radiography. In: Krupinski E. A, ed. *Medical Imaging 2000: Image Perception and Performance*. **3981**; 2000. pp. 135–44.
 19. Potter MC, Wyble B, Pandav R, Olejarczyk J. Picture detection in RSVP: features or identity? *J Exp Psychol Hum Percept Perform* 2011; **36**: 1486–94.
 20. Malsburg Cvonder, Von der Malsburg C. Binding in models of perception and brain function. *Current Opinion in Neurobiology* 1995; **5**: 520–6. doi: [https://doi.org/10.1016/0959-4388\(95\)80014-X](https://doi.org/10.1016/0959-4388(95)80014-X)
 21. Wolfe JM, Cave KR. The psychophysical evidence for a binding problem in human vision. *Neuron* 1999; **24**: 11–17. doi: [https://doi.org/10.1016/S0896-6273\(00\)80818-1](https://doi.org/10.1016/S0896-6273(00)80818-1)
 22. Treisman A. How the deployment of attention determines what we see. *Visual Cognition* 2006; **14**(4-8): 411–43. doi: <https://doi.org/10.1080/13506280500195250>
 23. Wolfe JM. Visual search in continuous, naturalistic stimuli. *Vision Research* 1994; **34**: 1187–95. doi: [https://doi.org/10.1016/0042-6989\(94\)90300-X](https://doi.org/10.1016/0042-6989(94)90300-X)
 24. Wolfe JM. Moving towards solutions to some enduring controversies in visual search. *Trends in Cognitive Sciences* 2003; **7**: 70–6. doi: [https://doi.org/10.1016/S1364-6613\(02\)00024-4](https://doi.org/10.1016/S1364-6613(02)00024-4)
 25. Evans KK, Treisman A. Perception of objects in natural scenes: is it really attention free? *J Exp Psychol Hum Percept Perform* 2005; **31**: 1476–92. doi: <https://doi.org/10.1037/0096-1523.31.6.1476>
 26. Biederman I. Perceiving real-world scenes. *Science* 1972; **177**: 77–80. doi: <https://doi.org/10.1126/science.177.4043.77>
 27. Oliva A. GIST of the scene. *Neurobiology of attention* 2005; **696**: 251–6.
 28. Potter MC, Faulconer BA. Time to understand pictures and words. *Nature* 1975; **253**: 437–8. doi: <https://doi.org/10.1038/253437a0>
 29. Alvarez GA. Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences* 2011; **15**: 122–31. doi: <https://doi.org/10.1016/j.tics.2011.01.003>
 30. Greene M, Oliva A. Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cognitive Psychology* 2009; **58**: 137–76. doi: <https://doi.org/10.1016/j.cogpsych.2008.06.001>
 31. Evans KK, Horowitz TS, Wolfe JM. When categories collide: accumulation of information about multiple categories in rapid scene perception. *Psychol Sci* 2011; **22**: 739–46. doi: <https://doi.org/10.1177/0956797611407930>
 32. Oliva A, Torralba A. Modeling the shape of the scene: a holistic representation of the spatial envelope. *International Journal of Computer Vision* 2001; **42**: 145–75. doi: <https://doi.org/10.1023/A:1011139631724>
 33. Keller BM, Oustimov A, Wang Y, Chen J, Acciavatti RJ, Zheng Y, et al. Parenchymal texture analysis in digital mammography: robust texture feature identification and equivalence across devices. *J Med Imaging* 2015; **2**: 024501. doi: <https://doi.org/10.1117/1.JMI.2.2.024501>
 34. Berman D, Golomb JD, Walther DB. Scene content is predominantly conveyed by high spatial frequencies in scene-selective visual cortex. *Plos One* 2017; **12**: e0189828. doi: <https://doi.org/10.1371/journal.pone.0189828>
 35. Li H, Giger ML, Sun C, Ponsukcharoen U, Huo D, Lan L, et al. Pilot study demonstrating potential association between breast cancer image-based risk phenotypes and genomic biomarkers. *Med Phys* 2014; **41**: 031917. doi: <https://doi.org/10.1118/1.4865811>
 36. Gierach GL, Li H, Loud JT, Greene MH, Chow CK, Lan L, et al. Relationships between computer-extracted mammographic texture pattern features and BRCA1/2 mutation status: a cross-sectional study. *Breast Cancer Res* 2014; **16**: 424. doi: <https://doi.org/10.1186/s13058-014-0424-8>
 37. Evans KK, Birdwell RL, Wolfe JM. If you don't find it often, you often don't find it: Why some cancers are missed in breast cancer screening. *PLoS ONE* 2013; **8**: e64366. doi: <https://doi.org/10.1371/journal.pone.0064366>
 38. Kallenberg M, Petersen K, Nielsen M, Ng AY, Diao P, Igel C, et al. Unsupervised deep learning applied to breast density segmentation and mammographic risk scoring. *IEEE Trans. Med. Imaging* 2016; **35**: 1322–31. doi: <https://doi.org/10.1109/TMI.2016.2532122>
 39. Huynh BQ, Li H, Giger ML. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *J Med Imaging* 2016; **3**: 034501. doi: <https://doi.org/10.1117/1.JMI.3.3.034501>