**Proceedings Paper:**

Gandomkar, Ziba, Ekpo, Ernest U., Lewis, Sarah J. et al. (5 more authors) (2018) Detection of the abnormal GIST in the prior mammograms even with no overt sign of breast cancer. In: 14th International Workshop on Breast Imaging (IWBI 2018). 14th International Workshop on Breast Imaging (IWBI 2018), 08-11 Jul 2018 SPIE , USA

https://doi.org/10.1117/12.2318321

# Does the strength of the Gist signal predict the difficulty of breast cancer detection in usual presentation and reporting mechanisms?

Ziba Gandomkar[*a], Ernest U. Ekpo[a], Sarah J. Lewis[a], Karla K. Evans[b], Kriscia Tapia[a], PhuongDung Trieu[a], Jeremy M. Wolfe[c], Patrick C. Brennan[a]

[a] Image Optimisation and Perception, Discipline of Medical Imaging and Radiation Sciences, Faculty of Health Sciences, University of Sydney, Sydney, NSW 2006, Australia.

[b] Department of Psychology, University of York, Heslington, York, UK.

[c] Visual Attention Lab, Harvard Medical School, Cambridge, MA, USA.

## Abstract

This study measured the correlation between the strength of a Gist signal and case difficulty based on standard presentation and reporting mechanisms for 80 cases. Half of the cases contained biopsy-proven cancer while the remainder were normal and confirmed to be cancer-free for at least two years of follow-up. In the Gist experiment, seventeen breast radiologists and physicians gave an abnormality score on a scale from 0 (confident normal) to 100 (confident abnormal) to unilateral CC mammograms following a very brief, 500 millisecond presentation of the image. Independently, each mammogram was assessed by a separate sample of at least 40 radiologists using standard presentation and reporting mechanisms, with these readers asked to locate any cancers present. All readers reported at least 1000 cases annually. For each case and each category, the percentage of correct reports served as an objective measure of case difficulty. For each of the 17 readers, the association between the abnormality scores from the gist study and detection rates from the earlier reports was examined using Spearman correlation. None of the coefficients were significantly different from zero ($p>0.05$). For the normal cases, the correlation coefficient between abnormality scores and detection rates for the 17 readers ranged from -0.262 to 0.258, and for cancer -0.180 to 0.309. The results suggest that the gist signal may indicate the presence of cancer, using mechanisms other than those employed in usual reporting, and might be exploited to improve breast cancer detection.

**Keywords**: Breast cancer, Breast cancer risk, Gist, Mammography, Prior mammograms.

## 1. Introduction

Early eye tracking experiments indicated that the scanpaths of both experienced and radiologists and radiology trainees exhibited inconsistency with an orderly approach for visual search instructed in radiology textbook [1] . More than four decades ago, it was also shown that experienced radiologists can identify abnormal chest x-ray based on a half-second image presentation prior to detailed fixations at any location [2]. Kundel and Nodine suggested the global-focal search model for medical image interpretation to explain these findings [3]. The model suggests that gist or "gestalt" information, which is immediately perceived by an observer after image onset using low resolution peripheral vision, constrains subsequent fixations. Radiologists identify image perturbations using the gist information and then gaze is guided to the location of perturbation so that they can analyze and further investigate the area by using the high-resolution foveal vision. The global-focal model originally developed to explain the visual search for detecting pulmonary nodules and later expanded to the mammographic visual search [4, 5].

Recent studies of the "gist" signal obtained from the mammographic images suggested that the gist might not necessarily provide information regarding the location of abnormality. In [6], it was shown that although the radiologists can detect abnormal cases at above-chance levels after a half-second presentation of mammograms, their performance is at chance-level in localizing the abnormality localization. It was also shown that experts can differentiate abnormal cases from normal ones based on a half-second presentation of normal mammograms contralateral to malignancies. These mammograms did not contain any overt sign of cancer [7]. However, women who have cancer in one breast are more likely to have future cancer in the contralateral breast [8]. Therefore, the gist has a might have an ability to predict future cancer. Moreover, using prior images that contained no overt evidence of cancer following detailed examination, readers could distinguish patients who would be eventually diagnosed with cancer, from individuals who would not develop the disease with a discriminatory power of up to 0.65 based on a half-second glimpse of the mammogram [9]. These two experiments show the possible value of the second aspect of Gist, described above, that suggests that global statistics, from across the entire image contain information about cancer risk.

In the present study, we explored whether there is any correlation between the strength of the Gist signal and the case difficulty assessed by the usual presentation and reporting mechanisms. We hypothesize that the gist signal is a cancer indicator independent from the information on which radiologists rely in usual presentation and reporting mechanism.

## 2. Materials and Methods

The experimental procedure is summarized in Figure 1. For the purpose of this study, we required (i) the radiologist's Gist response and (ii) a measure for case difficulty following usual image presentation and reporting mechanisms. As suggested in [10], the error rates (ratio of cases that were interpreted incorrectly) in usual reporting mechanism can serve as an objective measure for case difficulty. While the actual error rate is considered to be an ultimate measure for assessing case difficulty, many studies have had to use subjective measures (e.g. difficulty ratings by experts) because it can be challenging to recruit the large number of readers needed for error rate calculations. However, in our study, thanks to the BreastScreen Reader Assessment Strategy (BREAST) platform [11], we had a unique opportunity to determine the case difficulty based on the actual error rates. Two experiments will be further discussed in this section.

### *2.1. Case set*

We randomly retrieved 80 cases from the BREAST case sets. Half of the cases contained a biopsy-proven cancer while the others were normal and confirmed to be cancer-free for at least 2 years of follow-up. Number of cancer cases with BI-RADS I-IV density categories were 4, 17, 19, and 0, respectively and number of normal cases in each one of the BI-RADS density categories were 8, 17, 14, and 1, respectively. The mean lesion size was 10.62 $mm^2$ and it ranged from 3 to 26 $mm^2$. Lesion distance from the center of display was on average 1513 pixels. Further information with regard to the lesion locations and types is shown in Table 1.

### *2.2. Gist Experiment*

In the Gist experiment, 17 radiologists and breast physicians gave an abnormality score on a scale from 0 (confident normal) to 100 (confident abnormal) to 80 unilateral CC mammograms following a 500 milliseconds presentation of the image. Participants were recruited at the Royal Australian and New Zealand College of Radiologists (RANZCR) Breast Imaging Group (BIG) general meeting in 2017, and all had expertise in breast imaging. All readers reported at least 1000 cases annually.

For the presentation of images, in-house MATLAB-based (Mathworks, Natick, MA, USA) computer application was used. The experimental procedure is shown in Figure 1. First, a cross was showed in the center of the screen for a half-second, then a unilateral CC mammogram was presented for a half-second. After this, a white breast mask corresponding to the mammogram was presented for a half-second to ensure that the reader stops processing the image. Finally, a rating interface was prompted and the reader was asked to rate whether the presented mammogram was normal or abnormal. Further information with regard to the experimental procedure can be found in [12]

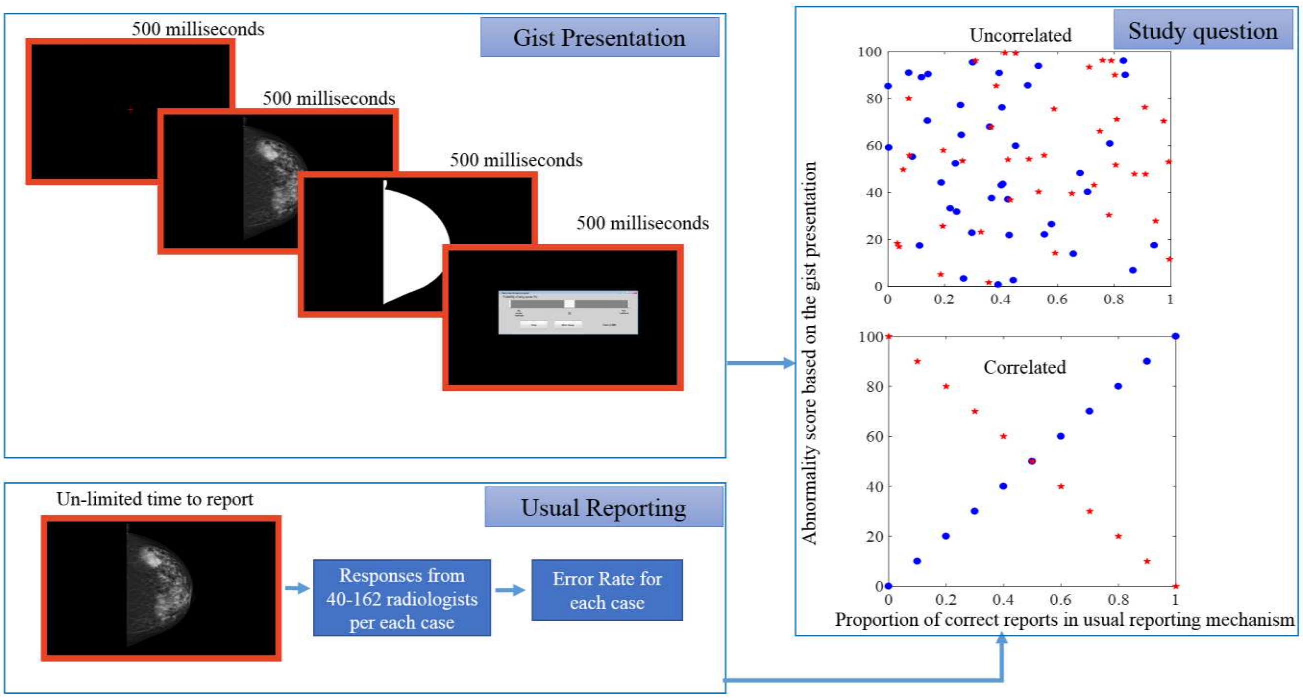


*Figure 1- Summary of the experiment; the blue and red circles in the correlation graphs represent the cancer and normal cases. In this study we examined the association between the abnormality scores from the gist study and detection rates from the usual reporting.*

Table 1- The characteristics of included cancer cases

| Location | | Lesion type | |
|---|---|---|---|
| Central | 8 | Architectural Distortion | 2 |
| Inner | 13 | Calcification | 5 |
| Outer | 18 | Discrete Mass | 6 |
| Retro areolar | 1 | Non-specific density | 11 |
| | | Speculated Mass | 5 |
| | | Stellate | 11 |

The MATLAB-based application was run on a Microsoft Surface Pro 4, which was connected to a Philips 28-inch LED (Model 288P6LJEB) monitor with a typical brightness of 300 cd/m² and a resolution of 3840 × 2160 pixels. All mammograms had an original resolution of 4915 × 5355 pixels and were resized prior to the experiment so that their heights were 1944 pixels.

### *2.3. Usual reporting mechanism*

Through BREAST [11], over 1000 radiologists either at scientific meetings or on-line have registered to do our test sets. Figure 2 shows a snapshot of the BREAST software, which is used by radiologists to locate the cancer on mammogram, assign the confidence score, and select the lesion type. Here we ignored whether the selected lesion types were correct in considering a reportas a "correct" one. Radiologists initially interpret the images using two 5-mega pixel diagnostic monitors and use the BREAST software only for reporting. They were also asked to assign a confidence score on the scale of 2-5. Here, in line with what has be done in the BREAST software for calculating the readers' sensitivity and specificity, all lesions categorized as 3 or above were considered as cancer-containing annotations. For each of the selected cases, the BREAST archive of the responses was searched to extract all readers who assessed that particular case. Based on the readers' report, we calculated the proportion of correct reports for each case (i.e. 1-error rate). The number of readers who assessed each case varied from 40 to 162.

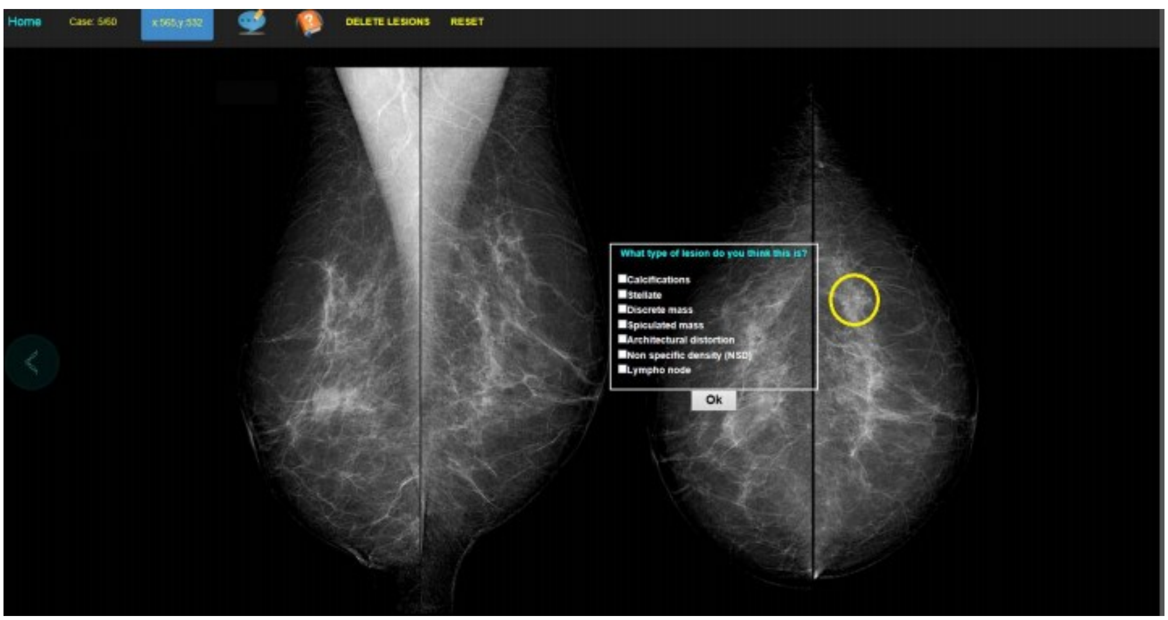


*Figure 2- The BREAST platform. Radiologists assessed the images using two 5-mega pixel diagnostic monitors and then reported the location of lesion on the images presented in the BREAST software. They were also asked to identify the lesion type and their confidence scores on the scale of 2-5. Here all lesions categorized as 3 or above were considered as areas annotated as malignant by an observer. The annotations were considered true-positive even if the selected lesion type did not match the correct classification of lesion type.*

### *2.4. Analysis*

If the Gist signal and the process underling the usual presentation and reporting mechanism are based on similar image information, the proportion of correct reports in the usual reporting of cancer cases should be correlated positively to the abnormality score based from the Gist experiment. For the normal cases, this correlation should be negative. On the other hand, an uncorrelated trend between these variables supports the hypothesis that the gist signal may indicate the presence of cancer, using mechanisms other than that employed in usual reporting of the case. These two possible trends are shown in Figure 1.

For each of the 17 readers, the association between the abnormality scores from the gist study and detection rates from the earlier reports was examined using Spearman correlation and the Hilbert-Schmidt Independence Criterion (HSIC) [13]. The HSIC is a kernel dependence measure to quantify the dependence between two random variables. Experimental results showed that effectiveness of HSIC-based tests to determine dependency of random variables which were dependent but uncorrelated [13, 14]. We used bootstrap resampling on the aggregated data to obtain the HSIC test threshold as describe in [13].

## 3.Results

Table 2 summarized the characteristics of radiologists participated in the gist experiment and their performance in categorization of cases as normal or cancer. For the cancer cases the correlation coefficient between abnormality scores and detection rates for the 17 readers ranged -0.180 to 0.309. For the normal cases, the correlation coefficient values for the readers varied between -0.262 and 0.258. None of the coefficients were significantly different from zero ($p>0.05$).

The HSIC test showed that for all readers, the case difficulty in usual presentation of mammograms was independent from the abnormality score given by all readers. Figure 3 shows the average gist score against the proportion of correct reports. In this figure, we calculated the average Gist signal by taking the mean of abnormality scores given by all 17 readers in the Gist experiment for each image.

Table 2- The characteristics of radiologists in the gist experiment and their performances as measured by the area under receiver operating curve (AUC).

| | Years working as screen reader | Hours reading mammogram per week | Number of mammograms per week | Years reading mammograms | AUC |
|---|---|---|---|---|---|
| **1** | 2 | <4 | 60-100 | 7 | 0.85 |
| **2** | 0 | 5 to 10 | 20-59 | 4 | 0.824 |
| **3** | 17 | 5-10 | 151-200 | 20 | 0.829 |
| **4** | 13 | 5-10 | >200 | 30 | 0.803 |
| **5** | 23 | 5-10 | >200 | 25 | 0.813 |
| **6** | 14 | 5-10 | 151-200 | 17 | 0.84 |
| **7** | 24 | >30 | >200 | 25 | 0.779 |
| **8** | 28 | 5-10 | 100-150 | 29 | 0.769 |
| **9** | 8 | 5-10 | 151-200 | 8 | 0.869 |
| **10** | 20 | 5-10 | 60-100 | 30 | 0.825 |
| **11** | 17 | 5-10 | >200 | 20 | 0.918 |

| | | | | | |
|---|---|---|---|---|---|
| **12** | 1 | 5-10 | 60-100 | 2 | 0.815 |
| **13** | 10 | 16-20 | >200 | 15 | 0.746 |
| **14** | 7.5 | 5-10 | >200 | 7.5 | 0.654 |
| **15** | 12 | <4 | 60-100 | 13 | 0.733 |
| **16** | 5 | 5 to 10 | 101-150 | 10 years | 0.865 |
| **17** | 19 | 16-20 | >200 | 20 years | 0.791 |

## 4. Conclusion

Humans are capable of recognizing the meaning, or Gist, of a complex visual scene, in a single glance. Radiologists, as experts in visual search tasks, can identify abnormal cases based on a rapid presentation of an image. There are two main manifestations of a Gist signal in radiology. The first one proposed by Kundel and Nodine [1] suggests that gist or “gestalt” information constrains subsequent fixations and guides visual search to the region of the image containing the abnormality. On the other hand, Evans et al [2] described a gist signal that could be used to categorize images as normal or abnormal but that was unrelated to the location of a focal abnormality. In this study we investigated whether the gist signal is related to the case difficulty in usual presentation and reporting mechanism of mammograms.

The results suggest that the strength of the Gist signal of the abnormal is not associated with the difficulty of the case when reported on in normal circumstances. The gist of the abnormal was detected by most of the readers in some of the difficult cases, for which radiologists were prone to make errors during reporting. This suggests that gist response may contain information different from the usual presentation of image for identifying a cancer, which should not be ignored or overruled following more detailed image evaluation.

In this study, we used two different cohorts of radiologists to extract case difficulty and the gist responses. They were not matched based on their experience level, number of years in practice, volume of breast reading, etc. Also, number of readers in the gist experiment was limited (17 radiologists in the gist experiment versus 40-162 radiologists in usual reading experiment). A possible future work for this study could be investigating the performance of the same observer while assessing the same case set and exploring ways to combine the gist and standard readings of these images. Such a combination may achieve a higher overall performance, if the gist response provide information different from the usual reporting mechanism for identifying a malignancy.

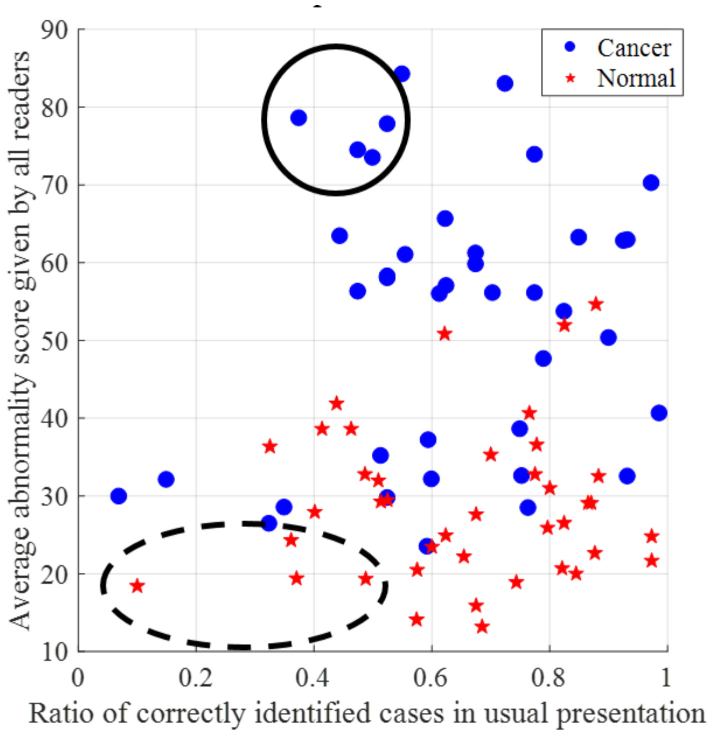


*Figure 3- The average abnormality score given by all readers against the ratio of correctly identified cases. Solid circle shows a few examples of cancer cases that exhibited a very obvious gist of the abnormal missed by more or about than 50% of readers in the usual presentation and reporting mechanism. Dashed oval shows a few examples of normal cases that were misidentified as cancer by majority of more or about than 50% readers in the usual presentation mechanism but had a very low gist signal.*

## Reference


[1] D. P. Carmody, H. L. Kundel, and L. C. Toto, "Comparison scans while reading chest images. Taught, but not practiced," Investigative Radiology, vol. 19, no. 5, pp. 462-466, 1984.

[2] H. L. Kundel and C. F. Nodine, "Interpreting chest radiographs without visual search," Radiology, vol. 116, no. 3, pp. 527-532, 1975.

[3] H. L. Kundel, C. F. Nodine, D. Thickman, and L. Toto, "Searching for lung nodules. A comparison of human performance with random and systematic scanning models," Investigative Radiology, vol. 22, no. 5, pp. 417-422, 1987.

[4] H. L. Kundel, C. F. Nodine, E. F. Conant, and S. P. Weinstein, "Holistic component of image perception in mammogram interpretation: gaze-tracking study," Radiology, vol. 242, no. 2, pp. 396-402, 2007.

[5] E. A. Krupinski, "Visual scanning patterns of radiologists searching mammograms," Academic radiology, vol. 3, no. 2, pp. 137-144, 1996.

[6] K. K. Evans, D. Georgian-Smith, R. Tambouret, R. L. Birdwell, and J. M. Wolfe, "The gist of the abnormal: Above-chance medical decision making in the blink of an eye," Psychonomic bulletin & review, vol. 20, no. 6, pp. 1170-1175, 2013.

[7] K. K. Evans, T. M. Haygood, J. Cooper, A.-M. Culpan, and J. M. Wolfe, "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, vol. 113, no. 37, pp. 10292-10297, 2016.

[8] X. Gao, S. G. Fisher, and B. Emami, "Risk of second primary cancer in the contralateral breast in women treated for early-stage breast cancer: a population-based study," International Journal of Radiation Oncology* Biology* Physics, vol. 56, no. 4, pp. 1038-1045, 2003.

[9] P. C. Brennan et al., "Radiologists can detect the 'gist'of breast cancer before any overt signs of cancer appear," Scientific reports, vol. 8, no. 1, p. 8717, 2018.

[10] M. A. Mazurowski, "Difficulty of mammographic cases in the context of resident training: preliminary experimental data," in Medical Imaging 2013: Image Perception, Observer Performance, and Technology Assessment, 2013, vol. 8673, p. 86730W: International Society for Optics and Photonics.

[11] P. Brennan, K. Tapia, J. Ryan, and W. Lee, "BREAST: a novel method to improve the diagnostic efficacy of mammography," in Medical Imaging 2013: Image Perception, Observer Performance, and Technology Assessment, 2013, vol. 8673, p. 867307: International Society for Optics and Photonics.

[12] Z. Gandomkar et al., "Detection of the abnormal gist in the prior mammograms even with no overt sign of breast cancer," in 14th International Workshop on Breast Imaging (IWBI 2018), 2018, vol. 10718, p. 1071804: International Society for Optics and Photonics.

[13] A. Gretton, K. Fukumizu, C. H. Teo, L. Song, B. Schölkopf, and A. J. Smola, "A kernel statistical test of independence," in Advances in neural information processing systems, 2008, pp. 585-592.

[14] K. Fukumizu, A. Gretton, X. Sun, and B. Schölkopf, "Kernel measures of conditional dependence," in Advances in neural information processing systems, 2008, pp. 489-496.