Computer-extracted global radiomic features can predict the radiologists' first impression about the abnormality of a screening mammogram

Somphone Siviengphanom (iD) **, MEd[1],*, Sarah J. Lewis, PhD[1], Patrick C. Brennan, PhD[1],**
Ziba Gandomkar, PhD[1]

[1]Medical Image Optimisation and Perception Group, Discipline of Medical Imaging Science, Sydney School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW 2006, Australia

*Corresponding author: Somphone Siviengphanom, MEd, Medical Image Optimisation and Perception Group, Discipline of Medical Imaging Science, Sydney School of Health Sciences, Faculty of Medicine and Health, The University of Sydney, Susan Wakil Health Building D18, Sydney, NSW, 2006, Australia (ssiv6387@uni.sydney.edu.au)

Abstract

Objective: Radiologists can detect the gist of abnormal based on their rapid initial impression on a mammogram (ie, global gist signal [GGS]). This study explores (1) whether global radiomic (ie, computer-extracted) features can predict the GGS; and if so, (ii) what features are the most important drivers of the signals.

Methods: The GGS of cases in two extreme conditions was considered: when observers detect a very strong gist (*high-gist*) and when the gist of abnormal was not/poorly perceived (*low-gist*). Gist signals/scores from 13 observers reading 4191 craniocaudal mammograms were collected. As gist is a noisy signal, the gist scores from all observers were averaged and assigned to each image. The *high-gist* and *low-gist* categories contained all images in the fourth and first quartiles, respectively. One hundred thirty handcrafted global radiomic features (GRFs) per mammogram were extracted and utilized to construct eight separate machine learning random forest classifiers (*All, Normal, Cancer, Prior-1, Prior-2, Missed, Prior-Visible*, and *Prior-Invisible*) for characterizing *high-gist* from *low-gist* images. The models were trained and validated using the 10-fold cross-validation approach. The models' performances were evaluated by the area under receiver operating characteristic curve (AUC). Important features for each model were identified through a scree test.

Results: The *Prior-Visible* model achieved the highest AUC of 0.84 followed by the *Prior-Invisible* (0.83), *Normal* (0.82), *Prior-1* (0.81), *All* (0.79), *Prior-2* (0.77), *Missed* (0.75), and *Cancer* model (0.69). *Cluster shade, standard deviation, skewness, kurtosis,* and *range* were identified to be the most important features.

Conclusions: Our findings suggest that GRFs can accurately classify *high-* from *low-gist* images.

Advances in knowledge: Global mammographic radiomic features can accurately predict *high-* from *low-gist* images with five features identified to be valuable in describing *high-gist* images. These are critical in providing better understanding of the mammographic image characteristics that drive the strength of the GGSs which could be exploited to advance breast cancer (BC) screening and risk prediction, enabling early detection and treatment of BC thereby further reducing BC-related deaths.

Keywords: radiomics; screening mammograms; gist; breast cancer; machine learning; global radiomic features; current mammograms; prior mammograms.

Introduction

Clinicians such as radiologists, the experts in the interpretation of medical images, often report that they have an initial impression about an abnormality of a medical image or its "gist" after momentarily glancing at an image. Several observer studies provided evidence suggesting that in less than a half-second, radiologists are capable of distinguishing abnormal mammograms[1] or chest X-rays from the normal ones.[2] A study focusing on mammographic images indicated that radiologists can distinguish abnormal from normal mammograms, containing no obvious localized sign of cancer, contralateral to the malignancies.[3] It was also shown that based on a half-second image presentation, the gist of the abnormal can be detected in prior normal breast images belonging to women who were later diagnosed with breast cancer (BC) (ie, missed cancers based on prior images from a previous screening round).[4-6] Moreover, previous studies showed that women with a false positive (FP) result on screening mammogram or who have had a cancer in one breast have elevated risk of

developing a future/contralateral BC.[7,8] These findings imply that the initial impression of a radiologist does not necessarily have a localized source within the image but is possibly driven by the mammographic textural features and not associated with other risk factors such as breast density.[9-12]

In medical image perception literatures, the holistic perceptual processing of radiological instantaneous view of an image's abnormality features is often called the gist signal, which has been extensively studied to better understand radiological errors and enhance screening performance.[3-6,11-21] From the human visual system, one may claim that gist contains both global and local source of information about "what" and "where" the abnormality in an image may be, correspondingly.[12,22,23] A large body of gist literature, studying gist on images without a localized cancer sign such as contralateral mammograms, revealed that gist is associated with global "what" information occurring without prior location of that information.[3-5,13-15,23,24] Views from the basic vision science domains suggest that gist information can be rapidly extracted (eg, using just as small as a

half-second of an image viewing time) from the entire global image properties in a "non-selective" manner, while identification of the information occurs from the local properties (ie, the "selective" pathway).[16,17,25,26] With that, gist is evidently a global rapid signal which will be referred to as the "global gist signal" (GGS) in this article.

Numerous gist studies focusing on the process of understanding natural scenes suggest that only after a brief sight of an image, observers are capable of categorizing these images into classes such as indoor or outdoor.[27,28] The human visual system relies on texture features in recognizing the global gist of the natural scenes,[29-31] demonstrating that texture analysis can provide strong holistic cues for quick scene recognition. Previous studies have shown that first and second order (Gray Level Co-occurrence Matrix [GLCM]-based) statistical features are relevant for texture analysis in typical viewing modes.[32-34] However, it is unclear whether these features are also relevant for the perceived GGS of a mammogram, which is extracted briefly after image onset based on a half-second image presentation. Our hypothesis is that a set of global radiomic features (GRFs), including both first order statistics (FOS) and GLCM-based, can differentiate between mammographic textures that result in strong gist signals (ie, *high-gist*) and those that result in weak ones (ie, *low-gist*). Therefore, this experimental study aims to (i) explore whether a set of GRFs comprising FOS and GLCM-based statistics extracted from mammograms can predict *high-* from *low-gist* images and (ii) identify the most important GRFs that drive the GGS. We focus only on these two extreme sides of the gist spectrum (highest and lowest gist images) to ensure that a clear difference between them at their maximum level is achieved. Moreover, mammographic images' textural features have been linked to the presence of BC[35,36] and elevated BC risk.[37] Therefore, we included images of normal cancer-free and cancer-containing cases, and high-risk individuals, and considered each class separately to ensure that our model is not capturing the signal related to the presence of BC or elevated BC risk, but rather capturing the gist signal.

In addition to enhancing our comprehension of the gist signal's nature when dealing with complex textures, the present study could yield insights into radiological errors in screening mammograms. Identifying features on normal images which signal a strong gist of the abnormal (FP) and missed cancer cases which signal a weak gist (false negative [FN]) could help in educating radiologists about possible appearances of FP and FN cases so that their diagnostic errors could be better monitored and drastically reduced. Since BC is the major cause of cancer-related death in females worldwide and the efficiency of mammographic screening interpretation depends largely on the visual expertise of clinician radiologists, errors such as FP and FN are significant.[38] While FP errors can trigger huge annual healthcare expenses of almost US$3 billion (an example from the United States) and an increased patient anxiety and depression, FN mistakes could result in increased treatment complications and reduced overall survival due to delayed detection of BC.[39,40] Hence, providing a better understanding about the underlying reasons for diagnostic errors in the interpretation of screening mammograms is crucially important.

Materials and methods

Institutional ethics approval of all experiment procedures was obtained from the Human Research Ethics Committee of the University of Sydney (protocol no. 2020/324). Informed consent from all subjects participated in the study was acquired prior to the data collection.

Participants

Seventeen Australian and New Zealand radiologists and breast physicians were invited to participate in this study. These readers were identified as the "gist expert", through their participation in the previous gist experiment[15] where they demonstrated superior ability to detect reliable and accurate gist of the abnormal in mammographic images. Full details of their selection and identification were described in a previous study[15] which also showed that the superior ability of these "gist experts" was not associated with factors such as their years of experience reading mammograms or number of mammograms interpreted each week. A total of 13 readers (11 radiologists and 2 breast physicians) were recruited.

Prior to the data collection, the 13 participants' characteristics and general workload details (Table 1) were collected via an online questionnaire. Majority of the participants were female (77%) radiologists (85%) working part-time (69%) as screen readers (92%) for the national screening program BreastScreen Australia and did not undertake a fellowship (62%) but had association with a university or educational institute (54%). At the time of participating in the gist experiment, on average, the participants spent 9.5 h per week reading mammograms and read 237 screening cases per week. With that, 70% of their time was dedicated to reading breast images, while 24.6% of their time was dedicated to reading diagnostic mammograms. The participants also had a median of 17.5 years being a certified BreastScreen reader, 21 years registered as a screen reader, and 21.5 years reading mammograms. Of those who had a fellowship training in mammography reading, the median duration of their training program was 6 months in which they completed it on average of 10 years ago. Lastly, on average, the participants performed 250 breast biopsy examinations in the last 12 months with a median of four times in a month correlating/reviewing radiology-pathology findings for biopsy cases, and three times in a month participating in a Multi-Disciplinary Meeting.

Mammographic images

For the purpose of conducting this gist experiment and radiomics analysis, mammographic images and binary (black and white) masks were required. A dataset of de-identified 4191 craniocaudal (CC) unilateral digital imaging and communications in medicine (DICOM) mammograms were obtained from the archive of BreastScreen Australia, which were acquired from asymptomatic women aged between 50 and 75. These images were obtained using a variety of vendors, that is, Siemens (Munich, Germany), Hologic (Hologic, Inc., Marlborough, MA, USA), Sectra (Sectra, Linköping, Sweden), Fujifilm (Fujifilm Corporation, Minato City, Tokyo, Japan), Philips (Philips Healthcare, Amsterdam, the Netherlands), and Konica Minolta (Marunouchi, Chiyoda, Tokyo, Japan). Next, binary masks were generated from 4191 DICOM images to extract breast area from its background using a standard thresholding gray level intensity value of 100. Due to computational efficiency reasons (to save computational resources, ie, time and memory), DICOM images and masks were then converted to TIFF

Table 1. Participants' characteristics and general workload details at the time of participating in the gist experiment ($n = 13$).

Characteristics			Count (%)
Gender (female, male)			10 (77%), 3 (23%)
Discipline (radiologist, breast physician)			11 (85%), 2 (15%)
Working full-time (yes, no)			4 (31%), 9 (69%)
Being a screen reader (yes, no)			12 (92%), 1 (8%)
Fellowship-trained in mammography reading (yes, no)			5 (38%), 8 (62%)
Whether affiliated with a university or educational institute (yes, no)			7 (54%), 6 (46%)
	Median	Min-Max	25th-75th percentile
No. of hours per week currently spent in reading mammograms	9.5	1.0-30.0	3.0-19.0
No. of screening cases read per week	237	5-800	150-400
Percentage of time dedicated to reading breast images	70	5-95	45-78
Percentage of time dedicated to reading diagnostic mammograms	24.6	0-90.2	9.0-35.0
No. of years certified as a BreastScreen reader	17.5	0.5-22	7.5-20
No. of years registered as a screen reader	21	2-30	12-25
No. of years reading mammograms	21.5	0-34	8.5-25
Duration of fellowship training in mammography reading (months)	6	6-12	6-12
No. of years since completing fellowship training in mammography reading	10	5-16	0-20.0
No. of breast biopsy examinations performed in the last 12 months	250	30-100	77-400
Frequency of correlating/reviewing radiology-pathology findings for biopsy cases in a month	4	1-100	4-40
Frequency of participating in a Multi-Disciplinary Meeting in a month	3	1-6	3-4
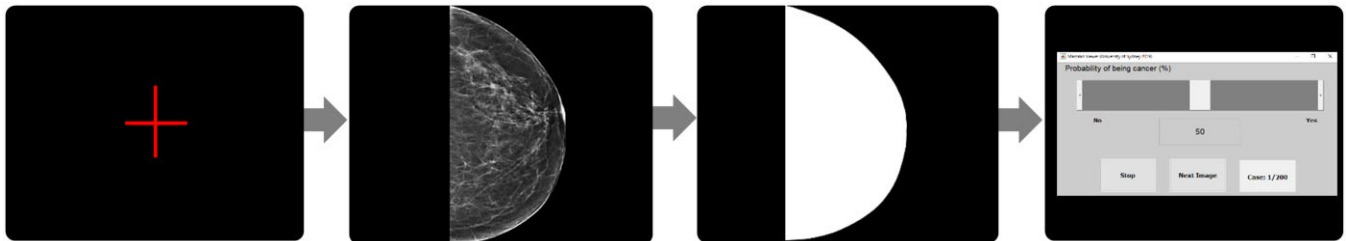
Figure 1. Gist experiment. For a half-second each, a red cross sign was firstly displayed in the centre of the screen, followed by a mammogram and then a corresponding white mask representing the breast area appeared to ensure conclusion of image visual processing. Lastly, without a time limit, a rating screen was shown to collect a gist score (ie, the probability of the image being abnormal) based on a scale of 0-100 (0 means completely confidence that the image was normal while 100 means absolute confidence that the image was abnormal).

format with all the right CC images being flipped to the left for consistent left side chest wall on all images. The TIFF images and binary masks were then cropped based on the maximum breast region size and used for the gist experiment (Fig. 1) and radiomics analysis (Fig. 2).

Since the gist of the abnormal has been reported to be presented in both current and prior images[3-6,11,13-16] and to include various levels of gist strength in our analysis, we included all the eight types of the mammograms in our dataset. The eight types were "Normal" ($n = 1409$), "Cancer" ($n = 923$), "Prior-1" ($n = 1438$), "Prior-2" ($n = 419$), "Prior-3" ($n = 2$), "Missed" ($n = 375$), "Prior-Visible" ($n = 308$), and "Prior-Invisible" ($n = 755$). The "Normal" category included current mammograms of women reported and confirmed to be cancer-free normal by at least two independent expert radiologists (over 20 years of experience) with negative follow up mammograms acquired 2 years after. The "Cancer" had current cancer images contained biopsy proven malignancy. The "Prior-1", "Prior-2", and "Prior-3" contained prior no visible cancer signs mammograms, obtained 2, 4, and 6 years before current cancer mammograms, correspondingly. The "Missed", "Prior-Visible", and "Prior-Invisible" were categorized by two expert radiologists or a third expert radiologist as an arbitrator if disagreement occurred between the two expert radiologists,[41] using "Prior-1" category images. These three categories, respectively,

included prior mammograms with actional (ie, recall), non-actional (ie, no recall), and no visible cancer signs that were reported as normal, but a later screen showed a biopsy-proven cancer.

Collecting gist scores

To collect the gist signals/scores from the 13 participants, a previously developed multi-observer gist experimental protocol[4] was used. Participants were asked to assess the 4191 CC images in 11 batches (381 images per batch selected randomly) and provide the probability of the image being abnormal/gist score based on a scale of 0 to 100 (0 = completely confidence the image was normal while 100 = absolute confidence that the image was abnormal). An extensive description of the data collection process (Fig. 1) was provided in a previous study.[42] Considering gist has low intra and inter-observer variability, the collected gist scores from the 13 radiologists (who were experts and consistent with each other in their gist scores) were then averaged and used for the analysis, obtaining one gist score per image and eliminating the noise of the gist signal.[11,14]

Categorizing high- vs low-gist images

To identify *high-* vs *low-gist* images, the averaged gist score per image was used to classify images into *high-* and *low-gist* based on the 75th and 25th percentiles of the images
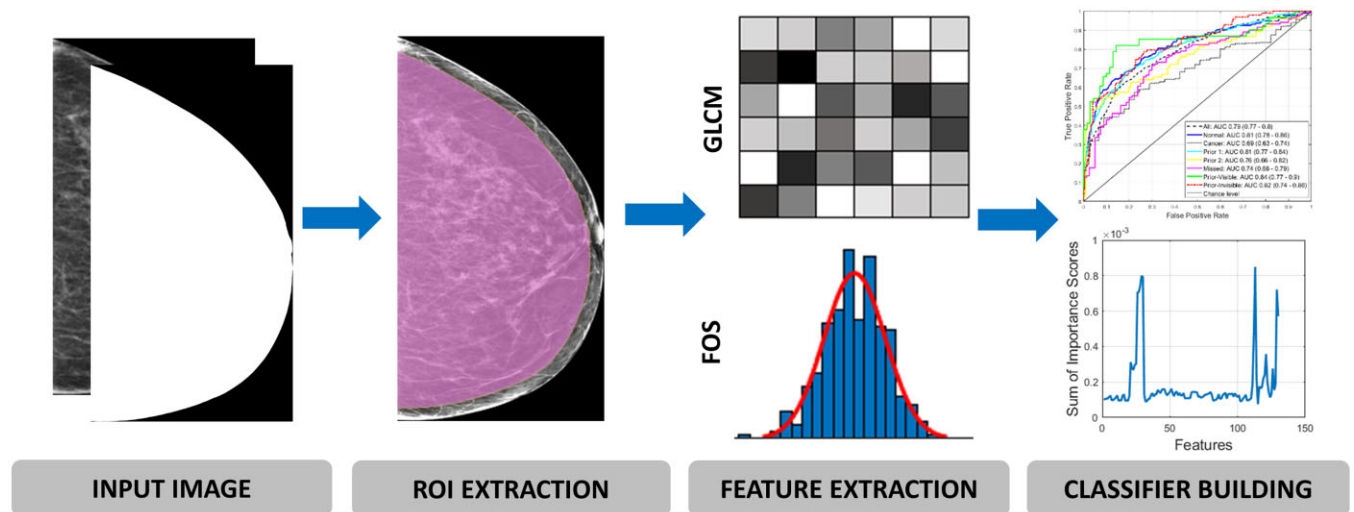
Figure 2. Radiomics analysis workflow. The first step was to obtain input images (unilateral craniocaudal mammograms and masks). Next, a ROI (pink region) was identified using MATLAB's image erosion algorithm to omit skin-air region. A set of 130 (110 GLCM-based and 20 FOS-based) global radiomic features per image were then computed and finally used to build eight machine learning classification models for predicting *high-gist* from *low-gist* images. Abbreviations: FOS = first order statistics; GLCM = gray level co-occurrence matrix; ROI = region of interest.

containing the highest and lowest gist scores, correspondingly. This gave a total of 1048 *high-* and 1048 *low-gist* images (namely "All" image type, including "Prior-3" images which had insufficient number of images, $n = 2$, to form its own group), 162 *high-* and 497 *low-gist* for "Normal", 545 *high-* and 71 *low-gist* for "Cancer", 268 *high-* and 370 *low-gist* for "Prior-1", 73 *high-* and 110 *low-gist* for "Prior-2", 117 *high-* and 80 *low-gist* for "Missed", 59 *high-* and 68 *low-gist* for "Prior-Visible", 92 *high-* and 222 *low-gist* for "Prior-Invisible" image type (Table 2). To accentuate an obvious difference between the highest and lowest gist images, only images in these upper and lower quartiles were used in the analysis.

Radiomics and statistical analysis

Using the 4191 images and masks, a set of 130 GRFs per image was extracted using the handcrafted features approach[43] with our in-house MATLAB-based programs. This is based on the region of interest identified through MATLAB's binary erosion algorithm using disk-shaped structuring element of 100 radius (distance from the origin to the edge of the disk),[44-46] to exclude the vendor post-processed background skin-air region in the mammogram in order to obtain the true breast tissue region.[47] Erosion algorithm was chosen because it is one of most reliable methods of the mathematical morphological image processing which has been widely used for exacting ROI and eliminating the noises in an image.[46] These features included 110 GLCM-based texture and 20 FOS-based features which have been shown to be valuable descriptors of mammographic appearances in measuring the contrast values of spatial inter-relationships between neighbouring pixels[48-50] and the distribution of single pixel intensity values within the image region of interest, correspondingly[49,50] (Table 3).

To predict *high-* from *low-gist* images, the 130 GRFs, derived from the corresponding image types ("All", "Normal", "Cancer", "Prior-1", "Prior-2", "Missed", "Prior-Visible", and "Prior-Invisible"—Table 2), were then used to build eight separate machine learning (ML) random forest classifiers (*All*, *Normal*, *Cancer*, *Prior-1*, *Prior-2*, *Missed*,

Prior-Visible, and *Prior-Invisible*) based on the ensemble of 500 decision trees using bootstrap aggregation method.[51,52] The random forest with bootstrap aggregation technique was chosen because of its built-in feature selection function that can minimize feature-overfitting and imbalanced data issues (ie, oversampling unique observations from the minority class to balance the data), and automatic estimation of feature importance for generating interpretable classifier.[53,54] Useful features for each predictive model were also recognized through feature important analysis of their importance scores estimated using MATLAB's predictor importance algorithm. To determine the useful GRFs based their total importance scores from each model, a scree test of exploratory factor analysis[55] was utilized. Importance scores show the usefulness of each feature in constructing the decision trees within the model, with larger values mean more impact on the predictions made by the model.

In evaluating the performance of the models, we trained and validated each model using the 10-fold cross-validation approach,[53,56] a reliable, unbiased, and accurate validation method for estimating the model's generalization performance. This involved randomly splitting the dataset into 10 groups, and while the first group was used once to test the predictive performance of the model, the rest of the groups were used to train the model. This process was repeated 10 times until each group was used once as a test set. The area under the receiver operating characteristic curve (AUC) was then used to evaluate the overall performance of the models for distinguishing *high-* from *low-gist* images.

MATLAB R2022a (MathWorks, Natick, MA, USA) was used to complete all radiomics and statistical analysis.

Results

GRFs predicting high- vs low-gist images

The overall performance of the eight models (*All*, *Normal*, *Cancer*, *Prior-1*, *Prior-2*, *Missed*, *Prior-Visible*, *Prior-Invisible*) is shown in Fig. 3. When differentiating *high-* from *low-gist* images, the *Prior-Visible* model reached the highest AUC of 0.84 (95% CI, 0.77-0.91) followed by the *Prior-*

Table 2. Image types included in the study ($n = 8$).

No.	Image type	Description	No. of high-gist images	No. of low-gist images	Total
1	All	Included all the image categories	1048	1048	2096
2	Normal	Current mammograms of women reported and confirmed to be cancer-free normal by at least two independent expert radiologists (20+ years of experience) with negative follow-up mammograms acquired two years after	162	497	659
3	Cancer	Current cancer mammograms of women contained biopsy proven malignancy	545	71	616
4	Prior-1	Prior mammograms with no visible cancer signs acquired two years before current cancer mammograms	268	370	638
5	Prior-2	Prior mammograms with no visible cancer signs acquired four years before current cancer mammograms	73	110	183
6	Missed	Prior-1 mammograms with actionable visible cancer signs (ie, recall) that were reported as normal, but a later screen showed a biopsy-proven cancer	117	80	197
7	Prior-visible	Prior-1 mammograms with non-actionable (ie, not recall) yet visible cancer signs that were reported as normal, but a later screen showed a biopsy-proven cancer	59	68	127
8	Prior-invisible	Prior-1 mammograms with no visible cancer signs that were reported as normal, but a later screen showed a biopsy-proven cancer	92	222	314

Invisible (0.83, 0.77-0.87), *Normal* (0.82, 0.79-0.86), *Prior-1* (0.81, 0.79-0.85), *All* (0.79, 0.77-0.81), *Prior-2* (0.77, 0.68-0.83), *Missed* (0.75, 0.70-0.81), and *Cancer* model (0.69, 0.65-0.75).

Important GRFs

Table 3 and Fig. 4 show the feature importance scores of the 130 GRFs derived from each of the eight models. When comparing each feature's importance score, ranked from high to low, in each individual model (Fig. 5), five features appear to be significant for most of the models. These features were *cluster shade* (feature no. 26-30, calculated across five orientations) of the GLCM,[48-50] *standard deviation* (feature no. 112), *skewness* (feature no. 113), *kurtosis* (feature no. 114), and *range* (feature no. 129 and 130, calculated across two dimensions) of the FOS.[49,50] *Cluster shade* is a measure of the symmetry distribution in an image. *Standard deviation* describes the variation or spreading of gray level intensity from the mean value. *Skewness* is a measure of the unevenness of the distribution. *Kurtosis* measures the extreme values of the distribution in the image. Also, *range* describes the difference between 95th and 5th percentile (feature no. 129), and 99th and 1st percentile of image gray level values (feature no. 130).

Discussion

For almost 50 years, research has shown that the holistic rapid visual impression of an image (ie, GGS) can convey a great amount of useful information about an image abnormality in the medical image interpretation and diagnostic process, aiming to significantly reduce FN and FP errors.[2,18,26] Recent medical image perception studies have recognized the existence of GGS shown to be valuable for providing information about the presence of BC to facilitate the improvement of early BC detection and future risk prediction.[3-6,13-16] However, little is known about which global computer extracted image features are correlated to the intensity of the GGS.[21] Here, for the first time, using radiomics approach,[43] we built upon the previous works and evaluated GRFs that drive the strength of the gist signals. We hypothesized that GRFs of *high-gist* mammograms are different from *low-gist* ones, based on the efficacy of radiomics in mammography[43] and the GGS, independent of lesion characteristics (eg, breast density), in identifying image abnormality.[3-5,11,13,14,16,24] We also examined which features are more important for distinguishing between *high-gist* mammograms from *low-gist* ones.

The results obtained from our eight ML classification models (*All*, *Normal*, *Cancer*, *Prior-1*, *Prior-2*, *Missed*, *Prior-Visible*, *Prior-Invisible*) demonstrated that the GRFs had the ability to accurately differentiate *high-* from *low-gist* images with as high as an AUC of 0.84 (Fig. 3). Interestingly, although the GGS has been shown to be helpful for predicting cancer images,[4,13] the *Cancer* model in our study achieved smaller AUC value of 0.69 in discriminating *high-* from *low-gist* cancer mammograms when compared to other seven models (AUCs of 0.75-0.84). This could be because of the large imbalanced dataset ($n = 474$) between the *high-gist* ($n = 545$) and *low-gist* ($n = 71$) images in the "Cancer" image category, making it difficult for the *Cancer* model to optimally learn the difference, thereby limiting its predictive capability. Nevertheless, our overall promising findings (ie, up to 0.84 AUC) indicate the significance of the GRFs in predicting *high-gist* images which could potentially be used to guide

Table 3. List of global radiomic features ($n = 130$) with feature importance scores.

No.	Feature name	Importance scores from each model							
		All	Normal	Cancer	Missed	Prior-Vis	Prior-Invis	Prior1	Prior2
1	Autocorrelation_1	0.0001	0.0003	0.0003	0.0010	0.0015	0.0005	0.0003	0.0008
2	Autocorrelation_3	0.0001	0.0003	0.0003	0.0009	0.0015	0.0006	0.0003	0.0009
3	Autocorrelation_5	0.0001	0.0003	0.0003	0.0008	0.0013	0.0006	0.0004	0.0009
4	Autocorrelation_9	0.0001	0.0003	0.0003	0.0009	0.0015	0.0005	0.0003	0.0007
5	Autocorrelation_11	0.0001	0.0003	0.0003	0.0008	0.0016	0.0006	0.0004	0.0009
6	Contrast_1	0.0001	0.0003	0.0003	0.0015	0.0007	0.0008	0.0003	0.0018
7	Contrast_3	0.0001	0.0002	0.0003	0.0013	0.0013	0.0005	0.0003	0.0012
8	Contrast_5	0.0001	0.0002	0.0002	0.0009	0.0018	0.0006	0.0003	0.0007
9	Contrast_9	0.0001	0.0002	0.0003	0.0009	0.0019	0.0007	0.0003	0.0008
10	Contrast_11	0.0001	0.0002	0.0003	0.0009	0.0021	0.0006	0.0004	0.0007
11	Correlation_m_1	0.0001	0.0003	0.0002	0.0009	0.0009	0.0007	0.0004	0.0017
12	Correlation_m_3	0.0001	0.0002	0.0003	0.0008	0.0013	0.0006	0.0003	0.0009
13	Correlation_m_5	0.0001	0.0002	0.0003	0.0010	0.0016	0.0005	0.0003	0.0008
14	Correlation_m_9	0.0001	0.0002	0.0002	0.0009	0.0016	0.0005	0.0003	0.0010
15	Correlation_m_11	0.0001	0.0002	0.0003	0.0013	0.0020	0.0006	0.0004	0.0007
16	Correlation_p_1	0.0001	0.0003	0.0003	0.0010	0.0008	0.0007	0.0004	0.0019
17	Correlation_p_3	0.0001	0.0002	0.0003	0.0007	0.0013	0.0006	0.0003	0.0009
18	Correlation_p_5	0.0001	0.0002	0.0003	0.0010	0.0014	0.0005	0.0003	0.0008
19	Correlation_p_9	0.0001	0.0002	0.0002	0.0009	0.0016	0.0005	0.0004	0.0010
20	Correlation_p_11	0.0001	0.0002	0.0003	0.0013	0.0020	0.0006	0.0004	0.0007
21	Cluster_prominence_1	0.0003	0.0011	0.0006	0.0051	0.0161	0.0027	0.0016	0.0043
22	Cluster_prominence_3	0.0003	0.0013	0.0006	0.0050	0.0136	0.0028	0.0015	0.0040
23	Cluster_prominence_5	0.0003	0.0013	0.0005	0.0055	0.0124	0.0026	0.0017	0.0048
24	Cluster_prominence_9	0.0003	0.0012	0.0005	0.0055	0.0139	0.0035	0.0018	0.0052
25	Cluster_prominence_11	0.0003	0.0013	0.0006	0.0050	0.0121	0.0029	0.0017	0.0044
26	Cluster_shade_1	0.0008*	0.0030*	0.0009*	0.0055	0.0249*	0.0071*	0.0035*	0.0119*
27	Cluster_shade_3	0.0008*	0.0033*	0.0010*	0.0051	0.0262*	0.0068*	0.0035*	0.0126*
28	Cluster_shade_5	0.0008*	0.0034*	0.0009*	0.0051	0.0277*	0.0075*	0.0036*	0.0138*
29	Cluster_shade_9	0.0009*	0.0038*	0.0009*	0.0066	0.0303*	0.0088*	0.0040*	0.0148*
30	Cluster_shade_11	0.0009*	0.0040*	0.0009*	0.0059	0.0266*	0.0092*	0.0039*	0.0119*
31	Dissimilarity_1	0.0001	0.0003	0.0003	0.0011	0.0008	0.0007	0.0003	0.0018
32	Dissimilarity_3	0.0001	0.0003	0.0002	0.0010	0.0008	0.0005	0.0003	0.0013
33	Dissimilarity_5	0.0001	0.0003	0.0002	0.0011	0.0015	0.0005	0.0003	0.0011
34	Dissimilarity_9	0.0001	0.0002	0.0003	0.0012	0.0011	0.0007	0.0003	0.0011
35	Dissimilarity_11	0.0001	0.0002	0.0003	0.0011	0.0012	0.0007	0.0004	0.0007
36	Energy_1	0.0001	0.0003	0.0002	0.0013	0.0012	0.0008	0.0004	0.0013
37	Energy_3	0.0001	0.0004	0.0003	0.0013	0.0011	0.0008	0.0004	0.0016
38	Energy_5	0.0001	0.0004	0.0003	0.0011	0.0013	0.0007	0.0003	0.0015
39	Energy_9	0.0001	0.0006	0.0003	0.0009	0.0019	0.0007	0.0003	0.0012
40	Energy_11	0.0001	0.0004	0.0002	0.0013	0.0016	0.0007	0.0004	0.0016
41	Entropy_1	0.0002	0.0003	0.0004	0.0012	0.0024	0.0010	0.0006	0.0012
42	Entropy_3	0.0002	0.0003	0.0004	0.0015	0.0017	0.0007	0.0004	0.0013
43	Entropy_5	0.0001	0.0003	0.0003	0.0015	0.0022	0.0008	0.0004	0.0011
44	Entropy_9	0.0002	0.0004	0.0004	0.0013	0.0018	0.0007	0.0004	0.0012
45	Entropy_11	0.0002	0.0004	0.0003	0.0015	0.0025	0.0008	0.0004	0.0017
46	Homogeneity_m_1	0.0002	0.0003	0.0004	0.0009	0.0021	0.0009	0.0004	0.0016
47	Homogeneity_m_3	0.0001	0.0003	0.0004	0.0013	0.0017	0.0006	0.0003	0.0011
48	Homogeneity_m_5	0.0001	0.0003	0.0003	0.0019	0.0019	0.0006	0.0004	0.0010
49	Homogeneity_m_9	0.0001	0.0003	0.0004	0.0017	0.0013	0.0006	0.0004	0.0015
50	Homogeneity_m_11	0.0001	0.0003	0.0003	0.0023	0.0013	0.0007	0.0004	0.0012
51	Homogeneity_1	0.0002	0.0003	0.0003	0.0008	0.0021	0.0009	0.0004	0.0019
52	Homogeneity_3	0.0001	0.0003	0.0004	0.0016	0.0018	0.0007	0.0003	0.0012
53	Homogeneity_5	0.0001	0.0003	0.0003	0.0018	0.0020	0.0007	0.0004	0.0010
54	Homogeneity_9	0.0002	0.0003	0.0004	0.0014	0.0018	0.0006	0.0004	0.0012
55	Homogeneity_11	0.0001	0.0003	0.0003	0.0025	0.0018	0.0006	0.0004	0.0011
56	Maximum_probability_1	0.0001	0.0003	0.0003	0.0013	0.0014	0.0010	0.0004	0.0013
57	Maximum_probability_3	0.0001	0.0004	0.0003	0.0013	0.0013	0.0008	0.0003	0.0015
58	Maximum_probability_5	0.0001	0.0004	0.0002	0.0011	0.0013	0.0008	0.0004	0.0014
59	Maximum_probability_9	0.0001	0.0005	0.0003	0.0011	0.0017	0.0007	0.0003	0.0012
60	Maximum_probability_11	0.0001	0.0004	0.0002	0.0012	0.0018	0.0007	0.0004	0.0014
61	Sum_of_sqaures_variance_1	0.0001	0.0003	0.0003	0.0009	0.0013	0.0005	0.0003	0.0009
62	Sum_of_sqaures_variance_3	0.0001	0.0003	0.0003	0.0009	0.0016	0.0005	0.0003	0.0010
63	Sum_of_sqaures_variance_5	0.0001	0.0003	0.0003	0.0009	0.0014	0.0005	0.0003	0.0009
64	Sum_of_sqaures_variance_9	0.0001	0.0003	0.0003	0.0009	0.0017	0.0004	0.0003	0.0009

(continued)

Table 3. (continued)

No.	Feature name	Importance scores from each model							
		All	Normal	Cancer	Missed	Prior-Vis	Prior-Invis	Prior1	Prior2
65	Sum_of_sqaures_variance_11	0.0001	0.0003	0.0003	0.0010	0.0015	0.0004	0.0003	0.0009
66	Sum_average_1	0.0001	0.0004	0.0003	0.0011	0.0012	0.0006	0.0003	0.0015
67	Sum_average_3	0.0001	0.0004	0.0004	0.0010	0.0013	0.0007	0.0004	0.0014
68	Sum_average_5	0.0001	0.0004	0.0004	0.0010	0.0014	0.0006	0.0004	0.0013
69	Sum_average_9	0.0001	0.0004	0.0004	0.0009	0.0013	0.0006	0.0004	0.0012
70	Sum_average_11	0.0001	0.0004	0.0003	0.0010	0.0017	0.0006	0.0003	0.0014
71	Sum_variance_1	0.0001	0.0003	0.0003	0.0009	0.0017	0.0006	0.0003	0.0010
72	Sum_variance_3	0.0001	0.0003	0.0003	0.0010	0.0015	0.0005	0.0003	0.0011
73	Sum_variance_5	0.0001	0.0003	0.0003	0.0011	0.0015	0.0006	0.0003	0.0011
74	Sum_variance_9	0.0001	0.0003	0.0003	0.0010	0.0017	0.0005	0.0004	0.0011
75	Sum_variance_11	0.0001	0.0003	0.0003	0.0011	0.0018	0.0006	0.0003	0.0012
76	Sum_entropy_1	0.0001	0.0006	0.0003	0.0015	0.0020	0.0009	0.0004	0.0017
77	Sum_entropy_3	0.0001	0.0006	0.0004	0.0015	0.0014	0.0007	0.0004	0.0019
78	Sum_entropy_5	0.0001	0.0006	0.0003	0.0017	0.0013	0.0010	0.0004	0.0014
79	Sum_entropy_9	0.0001	0.0006	0.0003	0.0018	0.0020	0.0007	0.0004	0.0015
80	Sum_entropy_11	0.0001	0.0006	0.0003	0.0017	0.0019	0.0008	0.0005	0.0020
81	Difference_variance_1	0.0001	0.0003	0.0003	0.0015	0.0008	0.0008	0.0003	0.0017
82	Difference_variance_3	0.0001	0.0002	0.0003	0.0015	0.0011	0.0005	0.0003	0.0011
83	Difference_variance_5	0.0001	0.0002	0.0002	0.0009	0.0018	0.0006	0.0004	0.0007
84	Difference_variance_9	0.0001	0.0002	0.0002	0.0010	0.0019	0.0006	0.0003	0.0009
85	Difference_variance_11	0.0001	0.0002	0.0003	0.0009	0.0021	0.0006	0.0004	0.0008
86	Difference_entropy _1	0.0001	0.0003	0.0003	0.0012	0.0013	0.0008	0.0004	0.0017
87	Difference_entropy _3	0.0001	0.0003	0.0003	0.0010	0.0010	0.0007	0.0003	0.0017
88	Difference_entropy _5	0.0001	0.0003	0.0002	0.0012	0.0012	0.0006	0.0004	0.0010
89	Difference_entropy _9	0.0001	0.0002	0.0003	0.0015	0.0011	0.0008	0.0004	0.0013
90	Difference_entropy _11	0.0001	0.0002	0.0003	0.0017	0.0011	0.0008	0.0005	0.0008
91	Information_measure_of_correlation1_1	0.0001	0.0003	0.0003	0.0011	0.0016	0.0010	0.0004	0.0028
92	Information_measure_of_correlation1_3	0.0001	0.0003	0.0003	0.0008	0.0018	0.0008	0.0003	0.0014
93	Information_measure_of_correlation1_5	0.0001	0.0003	0.0003	0.0011	0.0014	0.0007	0.0004	0.0011
94	Information_measure_of_correlation1_9	0.0001	0.0002	0.0003	0.0011	0.0009	0.0009	0.0004	0.0016
95	Information_measure_of_correlation1_11	0.0001	0.0003	0.0004	0.0017	0.0009	0.0007	0.0004	0.0008
96	Information_measure_of_correlation2_1	0.0001	0.0004	0.0003	0.0013	0.0017	0.0009	0.0004	0.0019
97	Information_measure_of_correlation2_3	0.0001	0.0004	0.0003	0.0011	0.0020	0.0008	0.0004	0.0015
98	Information_measure_of_correlation2_5	0.0001	0.0003	0.0003	0.0013	0.0017	0.0006	0.0004	0.0016
99	Information_measure_of_correlation2_9	0.0001	0.0002	0.0003	0.0012	0.0011	0.0009	0.0004	0.0017
100	Information_measure_of_correlation2_11	0.0001	0.0002	0.0003	0.0017	0.0014	0.0005	0.0004	0.0010
101	Inverse_difference_normalized _1	0.0001	0.0003	0.0003	0.0011	0.0009	0.0007	0.0004	0.0019
102	Inverse_difference_normalized _3	0.0001	0.0003	0.0002	0.0011	0.0011	0.0006	0.0003	0.0014
103	Inverse_difference_normalized _5	0.0001	0.0003	0.0003	0.0011	0.0016	0.0004	0.0003	0.0011
104	Inverse_difference_normalized _9	0.0001	0.0002	0.0003	0.0012	0.0015	0.0007	0.0004	0.0012
105	Inverse_difference_normalized _11	0.0001	0.0003	0.0003	0.0013	0.0011	0.0006	0.0004	0.0007
106	Inverse_difference_moment_ normalized _1	0.0001	0.0004	0.0003	0.0014	0.0008	0.0007	0.0003	0.0020
107	Inverse_difference_moment_ normalized _3	0.0001	0.0002	0.0003	0.0013	0.0013	0.0005	0.0003	0.0011
108	Inverse_difference_moment_ normalized _5	0.0001	0.0002	0.0002	0.0009	0.0015	0.0005	0.0003	0.0007
109	Inverse_difference_moment_ normalized _9	0.0001	0.0002	0.0003	0.0011	0.0019	0.0006	0.0003	0.0009

Table 3. (continued)

No.	Feature name	Importance scores from each model							
		All	Normal	Cancer	Missed	Prior-Vis	Prior-Invis	Prior1	Prior2
110	Inverse_difference_moment_ normalized _11	0.0001	0.0002	0.0003	0.0010	0.0019	0.0006	0.0004	0.0008
111	Mean	0.0002	0.0008	0.0004	0.0013	0.0019	0.0010	0.0005	0.0018
112	Standard_deviation	0.0005	0.0029*	0.0006	0.0085*	0.0262*	0.0050*	0.0025	0.0069
113	Skewness	0.0009*	0.0014	0.0006	0.0056	0.0114	0.0047*	0.0024	0.0065
114	Kurtosis	0.0002	0.0004	0.0008*	0.0019	0.0017	0.0014	0.0006	0.0012
115	Minimum	0.0001	0.0001	0.0003	0.0005	0.0004	0.0002	0.0001	0.0004
116	5th_percentile	0.0002	0.0007	0.0006	0.0016	0.0025	0.0009	0.0005	0.0009
117	10th_percentile	0.0002	0.0008	0.0004	0.0016	0.0015	0.0015	0.0006	0.0014
118	15th_percentile	0.0002	0.0007	0.0004	0.0021	0.0020	0.0021	0.0006	0.0017
119	20th_percentile	0.0002	0.0009	0.0004	0.0017	0.0026	0.0023	0.0007	0.0018
120	25th_percentile	0.0003	0.0011	0.0004	0.0017	0.0032	0.0020	0.0007	0.0024
121	Median	0.0004	0.0018	0.0003	0.0022	0.0027	0.0019	0.0010	0.0026
122	75th_percentile	0.0002	0.0008	0.0003	0.0018	0.0013	0.0015	0.0006	0.0011
123	80th_percentile	0.0002	0.0005	0.0003	0.0015	0.0016	0.0011	0.0004	0.0010
124	85th_percentile	0.0001	0.0004	0.0004	0.0016	0.0018	0.0010	0.0005	0.0013
125	90^{th}_percentile	0.0001	0.0006	0.0004	0.0024	0.0036	0.0013	0.0006	0.0015
126	95th_percentile	0.0003	0.0017	0.0005	0.0048	0.0094	0.0026	0.0012	0.0032
127	Maximum	0.0002	0.0003	0.0003	0.0013	0.0031	0.0007	0.0006	0.0010
128	Range_all (maximum less minimum)	0.0002	0.0004	0.0003	0.0017	0.0032	0.0011	0.0006	0.0021
129	Range5 (95th percentile less 5th percentile)	0.0008*	0.0017	0.0011*	0.0093*	0.0318*	0.0053*	0.0035*	0.0107*
130	Range2 (99th percentile less 1st percentile)	0.0006	0.0033*	0.0008*	0.0087*	0.0165	0.0060*	0.0025	0.0082

Feature nos. 1-110 are GLCM-based features[49-51] with feature parameters as 1, 3, 5, 9, and 11 being pixel distance between the pixel of interest and its neighbour. Feature nos. 111-130 are FOS-based features.[50,51]
* and highlighted in grey = important features based on their importance scores from the machine learning model. The larger the importance scores, the more important the features were for the model.
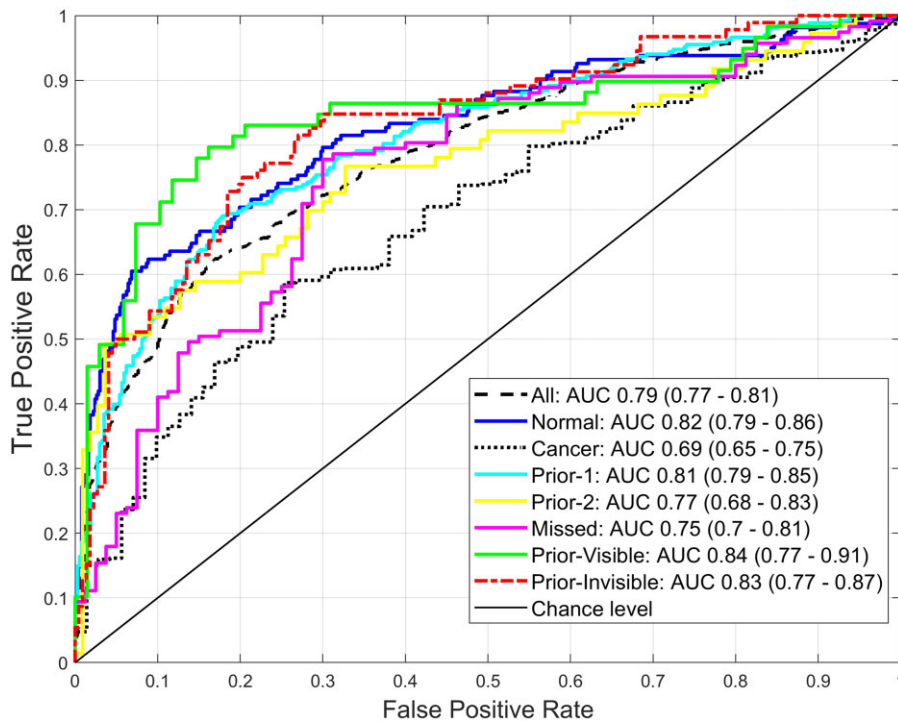Abbreviations: FOS = first order statistics; GLCM = gray level co-occurrence matrix.

Figure 3. Area under the receiver operating characteristic curve (AUC) for eight classifiers. When differentiating *high-* from *low-gist* images, all eight models using global radiomic features from each image category performed relatively well with the *Prior-Visible* images category model achieved the highest AUC of 0.84 (95% CI, 0.77-0.91) while the *Cancer* model had an AUC of 0.69 (95% CI, 0.65-0.75).
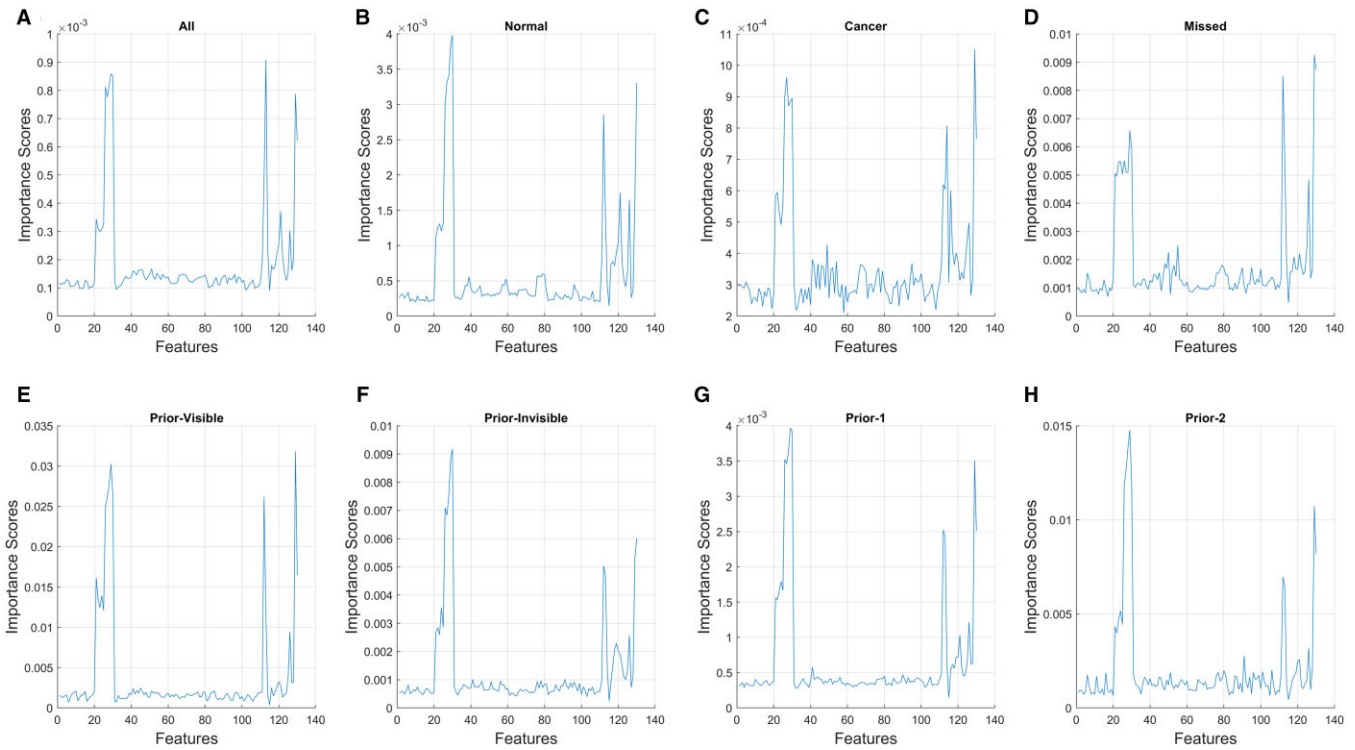
Figure 4. Line plot showing feature importance scores of the 130 features (no. 1-130) in each classifier (A)-(H). As shown, pattern of important features appears to be similar across the eight classifiers.
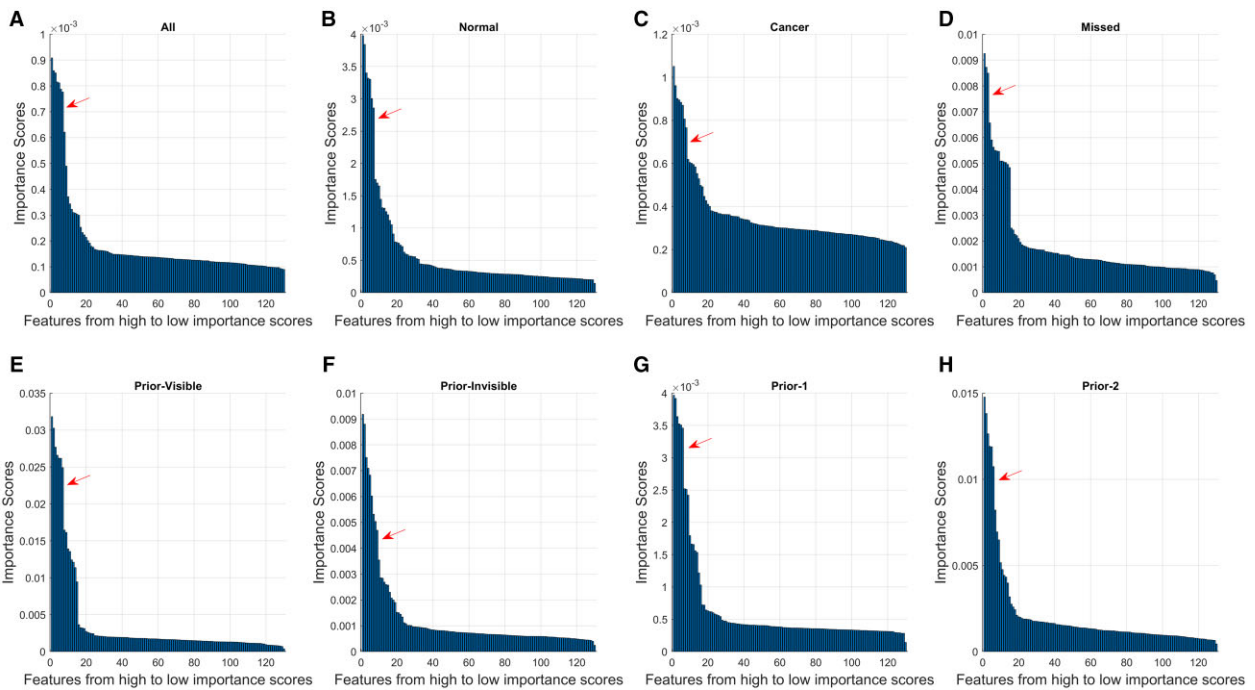
Figure 5. Scree bar chart for each of the eight classifiers showing important features (ranked from high to low importance scores) determined based on the identified steep slope (as shown in red arrow). Seven, seven, eight, three, seven, nine, six, and six most important features were identified for *All* (A), *Normal* (B), *Cancer* (C), *Missed* (D), *Prior-Visible* (E), *Prior-Invisible* (F), *Prior-1* (G), *Prior-2* (H) model, respectively.

the proper use of the GGS to support BC detection and prediction process.

Moreover, if the GGS can be considered as a BC risk factor, similar to breast density or genetic risk factor, our work may be applied as a supplemented artificial vision intelligence tool to signal clinician about an elevated risk of a positive cancer finding in mammographic screening.[3,5,13] It might also be used to augment the recent deep learning model[57] to further advance the accuracy of BC detection in screening mammography especially since the GGS has been shown to

be helpful in improving the performance of the deep learning model in detecting BC.[14] This could not only assist in lessening the possibility of FP and FN errors but also optimizing personalized screening strategy (eg, shorter screening interval) and increasing the chance of BC being detected early with improved therapy outcomes and survivorship. Conversely, more research efforts are needed to further evaluate and validate its clinical usefulness.

Interestingly, although studies[3,58] indicated that GGS of future cancerous abnormalities can be detected in fine details of the parenchymal texture of high-pass filtered mammograms, other works revealed that radiologists extract the GGS using low-resolution peripheral vision prior to detailed foveal processing of any image areas when separating abnormal from normal mammograms.[1,10,24] Our findings supported the relevance of the FOS and GLCM-based second-order textural features to the GGS, extracted by radiologists' peripheral vision. Prior studies[3,13] showed that observers failed to localize BC using such gist signal, indicating that the gist signal in the context of medical images is concordant with findings from basic vision sciences where observers struggle with tasks involving distinguishing phase differences[59] and show uncertainty in localizing bisection.[60] To move beyond qualitative evidence, a specific proposal for peripheral "texture processing" and its relevance to mammography are required. The current best theory for a model of capturing the appearance of textures is the Texture Tiling Model[61,62] which implicates the visual system with the summary statistics computed using the second-order statistics of primary visual cortex over local pooling regions.

In the context of basic vision science, the importance of texture features from non-medical images in identifying the global gist of natural scenes was discovered.[29-31] Our work based on breast medical images also showed similar results, in that, five useful features involving one GLCM-based textural and four FOS-based features (Table 3) from the ML models were found. In concordance with the suggestions from basic vision science studies,[33,34,63] features that are related to noticeable differences in the first order luminance, that is, *range* and *standard_deviation*, were among important features in the classification tasks. In particular, *range5* (feature no. 129) seems to be valuable in all the model except *Normal*, signifying that a high GGS in most image types may be motivated by the overall low range (difference between the 95th and 5th percentiles) of the image gray level value, resembling cancer characteristics. Similarly, among the second-order features, *cluster shade* (feature no. 26-30) exhibited importance in all tasks except for one, suggesting that the "uniformity of mammographic texture" could contribute to the radiologists' initial impression about the abnormality of a case. In this case, these features may be used as an indicator of whether a strong gist of the abnormal is present in an image, which could contribute to the early detection and risk prediction of BC. However, larger research is required to further examine the effectiveness of these features.

There were a few limitations in our study. First, this was the first exploratory study investigating the capability of handcrafted GRFs (ie, 110 GLCM-based texture and 20 FOS-based features) derived from 130 CC unilateral view mammograms[64] in predicting *high-gist* from *low-gist* images based on the gist scores of thirteen "gist experts".[15] The mediolateral oblique (MLO) view was not used in this study. While the entire images were viewed by the readers in the gist experiment, the background skin-air region of the images had to be excluded using erosion algorithm, because it has been enhanced by the vendors, to obtain the true breast region for GRF extraction. The results may be only applicable to the images and gist experts with characteristic/background (eg, being a screen reader with high screening workload) similar to the sample used in the study. Secondly, useful GRFs were determined based on their importance scores using a scree test[55] (Fig. 4), which may be less explicit for some models, for instance, the *Missed* model when compared to the *Normal* model. Also, only "Prior-1" images (attained two years prior to current cancer mammograms) were used to categorize "Missed", "Prior-Visible", and "Prior-Invisible" images. As with all model building, the output of the model is constrained by the parameters entered into the model. So, forthcoming studies should examine other prior mammograms acquired beyond 2 years before, MLO, raw[47] and high-pass filtered bilateral mammograms,[58] digital breast tomosynthesis images,[65] and images from other vendors (eg, GE, GE Healthcare, Chicago, IL, USA). These images may also contain valuable information to suggest an abnormality and/or an increased future malignancy risk, and therefore may result in different gist scores and GRFs. Also, gist experts from other location and using other morphological image processing methods (such as corrosion and closing[66]) may produce different results. Lastly, other potential important GRFs (eg, *cluster prominence*, feature no. 21-25) and types of GRFs (eg, deep learning features[11]) should also be analysed.

Conclusions

To summarize, this study suggests that a set of quantitative GRFs derived from mammographic images can accurately predict *high-* from *low-gist* images with five useful GRFs (one GLCM-based texture and four FOS-based features) discovered and emphasized. These findings are critically important in providing better understanding of the mammographic image characteristics that drive the strength of the GGS, which could be exploited to advance BC screening and risk prediction, enabling early detection and treatment of BC thereby further reducing BC-related deaths.

Acknowledgements

Author contributions

Conceptualisation, validation, and writing – review & editing, all authors. Data curation, formal analysis, investigation, methodology, and software, S.S and Z.G. Supervision and funding acquisition, Z.G., S.L., and P.B. Visualisation and writing – original draft, S.S.

Conflicts of interest

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this article. P.B. and Z.G., as well as an academic at the University of Sydney, serve as CEO and Co-founder (P.B.)/employee (Z.G.) of DetectedX, a company that aims to educate radiologists. The article will not impact on DetectedX's activities. All other authors of this article, other than being employee (S.L.)/student (S.S.) of the University of Sydney/employee (S.S.) of the University of Technology Sydney, declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Data availability

The dataset that supports the findings of this study is not publicly available as it contains proprietary information that the authors acquired through a license. Data are, however, available from the authors upon reasonable request.

Ethical approval

All procedures performed and reported in this article involving human participants were in accordance with the ethical standards of the University of Sydney's Human Research Ethics Committee [protocol no. 2020/324].

Consent

All individual participants in this study provided their written informed consent for participation and publication.

References

1. Gandomkar Z, Mello-Thoms C. Visual search in breast imaging. *Br J Radiol*. 2019;92(1102):20190057.
2. Kundel HL, Nodine CF. Interpreting chest radiographs without visual search. *Radiology*. 1975;116(3):527-532.
3. 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. *Proc Natl Acad Sci USA*. 2016;113(37):10292-10297.
4. Brennan PC, Gandomkar Z, Ekpo EU, et al. Radiologists can detect the 'gist' of breast cancer before any overt signs of cancer appear. *Sci Rep*. 2018;8(1):8717-8712.
5. Evans KK, Culpan 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(1099):20190136-20190110.
6. Gandomkar Z, Ekpo E, Lewis S, Evans K, Tapia K, Trieu P-D, et al. eds. Detection of the abnormal gist in the prior mammograms even with no overt sign of breast cancer. *SPIE 10718, 14th International Workshop on Breast Imaging (IWBI 2018)*; 2018; Atlanta, Georgia, United States: SPIE.
7. Henderson LM, Hubbard RA, Sprague BL, Zhu W, Kerlikowske K. Increased risk of developing breast cancer after a false-positive screening Mammogram. *Cancer Epidemiol Biomarkers Prev*. 2015;24(12):1882-1889.
8. Ramin C, Withrow DR, Davis Lynn BC, Gierach GL, Berrington De González A. Risk of contralateral breast cancer according to first breast cancer characteristics among women in the USA, 1992–2016. *Breast Cancer Res*. 2021;23(1):24.
9. Gastounioti A, Conant EF, Kontos D. Beyond breast density: a review on the advancing role of parenchymal texture analysis in breast cancer risk assessment. *Breast Cancer Res*. 2016;18(1):91-12.
10. Gandomkar Z, Ekpo E, Chen Z, Suleiman M, Li T, Tapia K, et al. eds. Expert radiologist performance does not appear to impact upon their capability in perceiving the gist of the abnormal on mammograms. *15th International Workshop on Breast Imaging (IWBI2020)*. SPIE; 2020:.
11. Gandomkar Z, Ekpo EU, Lewis SJ, Suleiman ME, Siviengphanom S, Li T, et al. eds. An end-to-end deep learning model can detect the gist of the abnormal in prior mammograms as perceived by experienced radiologists. *SPIE Medical Imaging 2021: Image Perception, Observer Performance, and Technology Assessment*. SPIE; 2021.
12. Kundel HL, Nodine CF, Conant EF, Weinstein SP. Holistic component of image perception in mammogram interpretation: gaze-tracking study. *Radiology*. 2007;242(2):396-402.
13. 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(6):1170-1175.
14. Gandomkar Z, Siviengphanom S, Ekpo EU, et al. Global processing provides malignancy evidence complementary to the information captured by humans or machines following detailed mammogram inspection. *Sci Rep*. 2021;11(1):20122-20112.
15. Gandomkar Z, Siviengphanom S, Suleiman Ma, Wong D, Reed W, Ekpo E, et al. eds. The reliability of radiologists' first impression interpreting a screening mammogram. *SPIE Medical Imaging 2022: Image Perception, Observer Performance, and Technology Assessment*; 4 April; San Diego, California, United States: SPIE; 2022.
16. Raat EM, Farr I, Wolfe JM, Evans KK. Comparable prediction of breast cancer risk from a glimpse or a first impression of a mammogram. *Cogn Res Princ Implic*. 2021;6(1):72.
17. Wolfe JM, Evans KK, Drew T. The first moments of medical image perception. In: E Samei, EA Krupinski, eds. *The Handbook of Medical Image Perception and Techniques*. 2nd ed. Cambridge University Press; 2018:188-196.
18. Waite S, Grigorian A, Alexander RG, et al. Analysis of perceptual expertise in radiology—current knowledge and a new perspective. *Front Hum Neurosci*. 2019;13:213.
19. Sheridan H, Reingold EM. The holistic processing account of visual expertise in medical image perception: a review. *Front Psychol*. 2017;8:1620.
20. Wolfe JM, Evans KK, Drew T, Aizenman A, Josephs E. How do radiologists use the human search engine? *Radiat Prot Dosimetry*. 2016;169(1-4):24-31.
21. Donovan T, Litchfield D, Crawford TJ. Editorial: medical image perception: how much do we understand it? *Front Psychol*. 2017; 8:2072.
22. Carrigan AJ, Wardle SG, Rich AN. Finding cancer in mammograms: if you know it's there, do you know where? *Cogn Res Princ Implic*. 2018;3(1):10-14.
23. Ungerleider L, Mishkin M. Two cortical visual systems. In: Ingle DJ, Goodale MA, Mansfield RJ, eds. *Analysis of Visual Behavior*. The MIT press; 1982:549-586.
24. Gandomkar Z, Ekpo E, Lewis S, Evans K, Tapia K, Trieu P, et al. eds. Does the strength of the gist signal predict the difficulty of

breast cancer detection in usual presentation and reporting mechanisms? *SPIE 10952 Medical Imaging 2019: Image Perception, Observer Performance, and Technology Assessment; 2019; San Diego, California, United States*; SPIE; 2019.

25. Wolfe JM, Võ ML, Evans KK, Greene MR. Visual search in scenes involves selective and nonselective pathways. *Trends Cogn Sci*. 2011;15(2):77-84.

26. Drew T, Evans K, Võ ML, Jacobson FL, Wolfe JM. Informatics in radiology: what can you see in a single glance and how might this guide visual search in medical images? *Radiographics*. 2013;33(1):263-274.

27. Joubert OR, Rousselet GA, Fize D, Fabre-Thorpe M. Processing scene context: fast categorization and object interference. *Vision Res*. 2007;47(26):3286-3297.

28. Rousselet G, Joubert O, Fabre-Thorpe M. How long to get to the "gist" of real-world natural scenes? *Visual Cognition*. 2005;12(6):852-877.

29. Renninger LW, Malik J. When is scene identification just texture recognition? *Vision Res*. 2004;44(19):2301-2311.

30. Oliva A, Torralba A. Modeling the Shape of the scene: a holistic representation of the spatial envelope. *Int J Comput Vision*. 2001;42(3):145-175.

31. Oliva A, Torralba A. Building the gist of a scene: the role of global image features in recognition. In: Martinez-Conde S, Macknik SL, Martinez LM, Alonso JM, Tse PU, editors. *Progress in Brain Research*. Vol. 155: Elsevier; 2006:23-36.

32. Nisbett WH, Kavuri A, Das M. On the correlation between second order texture features and human observer detection performance in digital images. *Sci Rep*. 2020;10(1):13510.

33. Chubb C, Nam J-H, Bindman DR, Sperling G. The three dimensions of human visual sensitivity to first-order contrast statistics. *Vision Res*. 2007;47(17):2237-2248.

34. Julesz B. Texture and visual perception. *Sci Am*. 1965;212(2):38-48.

35. Gandomkar Z, Tay K, Ryder W, Brennan PC, Mello-Thoms C. iCAP: an individualized model combining gaze parameters and image-based features to predict radiologists' decisions while reading mammograms. *IEEE Trans Med Imaging*. 2016;36(5):1066-1075.

36. Joseph S, Balakrishnan K, eds. Local binary patterns, haar wavelet features and haralick texture features for mammogram image classification using artificial neural networks. *Advances in Computing and Information Technology: First International Conference, ACITY 2011, Chennai, India, July 15-17, 2011 Proceedings*. Springer; 2011.

37. Gandomkar Z, Li T, Shao Z, Tang L, Xiao Q, Gu Y, et al. eds. Breast cancer risk prediction in Chinese women based on mammographic texture and a comprehensive set of epidemiologic factors. *Fifteenth International Workshop on Breast Imaging;Leuven, Belgium*. SPIE; 2020.

38. Krupinski EA. Current perspectives in medical image perception. *Atten Percept Psychophys*. 2010;72(5):1205-1217.

39. Ong M-S, Mandl KD. National expenditure for false-positive mammograms and breast cancer overdiagnoses estimated at $4 billion a year. *Health Aff (Millwood)*. 2015;34(4):576-583.

40. Li Y, Zhou Y, Mao F, et al. The influence on survival of delay in the treatment initiation of screening detected non-symptomatic breast cancer. *Sci Rep*. 2019;9(1):10158-10157.

41. Gandomkar Z, Lewis S, Siviengphanom S, et al. *Classification and Reviewing of Prior Screening Mammograms from Screen-Detected Breast Cancer Cases: SPIE*; 2022.

42. Gandomkar Z, Siviengphanom S, Suleiman MA, et al. Reliability of radiologists' first impression when interpreting a screening mammogram. *PLoS One*. 2023;18(4):e0284605.

43. Siviengphanom S, Gandomkar Z, Lewis SJ, Brennan PC. Mammography-based radiomics in breast cancer: a scoping review

44. TM Inc. imerode—Erode image: The MathWorks Inc.; 2023. Accessed October 17, 2023. https://au.mathworks.com/help/images/ref/imerode.html

45. TM Inc. strel—Morphological structuring element: The MathWorks Inc.; 2023. Accessed October 17, 2023. https://au.mathworks.com/help/images/ref/strel.html

46. Said KAM, Jambek AB, eds. *Analysis of Image Processing Using Morphological Erosion and Dilation. International Conference on Biomedical Engineering (ICoBE 2021);UniMAP, Malaysia (Virtual)*. IOP Publishing Ltd; 2021.

47. 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(8):4903-4917.

48. Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. *IEEE Trans Syst, Man, Cybern*. 1973;SMC-3(6):610-621.

49. Scapicchio C, Gabelloni M, Barucci A, Cioni D, Saba L, Neri E. A deep look into radiomics. *Radiol Med*. 2021;126(10):1296-1311.

50. Costaridou L. *Medical Image Analysis Methods*. Taylor & Francis; 2005.

51. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32.

52. Breiman L. Bagging predictors. *Mach Learn*. 1996;24(2):123-140.

53. Couronné R, Probst P, Boulesteix A-L. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinform*. 2018;19(1):270-214.

54. TM Inc. fitcensemble—Fit ensemble of learners for classification The MathWorks Inc.; 2023. Accessed July 28, 2023. https://au.mathworks.com/help/stats/fitcensemble.html

55. Ledesma RD, Valero-Mora P, Macbeth G. The Scree test and the number of factors: a dynamic graphics approach. *Span J Psychol*. 2015;18:E11.

56. Bengio Y, Grandvalet Y. No unbiased estimator of the variance of K-fold cross-validation. *The J Mach Learn Res*. 2004;5(2004):1089-1105.

57. Shen L, Margolies LR, Rothstein JH, Fluder E, McBride R, Sieh W. Deep learning to improve breast cancer detection on screening mammography. *Sci Rep*. 2019;9(1):12495-12412.

58. Raat EM, Evans KK. Early signs of cancer present in the fine detail of mammograms. *PLoS One*. 2023;18(4):e0282872.

59. Bennett PJ, Banks MS. The effects of contrast, spatial scale, and orientation on foveal and peripheral phase discrimination. *Vision Res*. 1991;31(10):1759-1786.

60. Levi DM, Klein SA. Sampling in spatial vision. *Nature*. 1986;320(6060):360-362.

61. Freeman J, Simoncelli EP. Metamers of the ventral stream. *Nat Neurosci*. 2011;14(9):1195-1201.

62. Rosenholtz R, Huang J, Raj A, Balas BJ, Ilie L. A summary statistic representation in peripheral vision explains visual search. *J Vis*. 2012;12(4):14.

63. Julesz B. Visual pattern discrimination. *IEEE Trans Inform Theory*. 1962;8(2):84-92.

64. Gandomkar Z, Ekpo EU, Lewis SJ, Evans KK, Tapi K, Li T, et al. eds. The strength of the gist of the abnormal in the unilateral and bilateral mammograms. *Medical Image Perception Society (MIPS) XVIII Conference; Salt Lake City, Utah, USA*. Medical Image Perception Society; 2019.

65. Wu CC, D'Ardenne NM, Nishikawa RM, Wolfe JM. Gist processing in digital breast tomosynthesis. *J Med Imaging (Bellingham)*. 2020;7(2):022403.

66. Zhang F, Wu Y, Xu M, Liu S, Peng C, Gao Z. A morphological image segmentation algorithm for circular overlapping cells. *Intell Autom Soft Comput*. 2022;32(1):301-321.