# Fractal Analysis of Visual Search Activity for Mass Detection During Mammographic Screening

Folami Alamudun[a], Hong-Jun Yoon[a], Kathleen B. Hudson[b], Garnetta Morin-Ducote[b], Tracy Hammond[c], Georgia D. Tourassi[a]

[a] Health Data Sciences Institute, Oak Ridge National Laboratory,

Oak Ridge, TN 37831;

[b] Department of Radiology, University of Tennessee Medical Center at Knoxville,

Knoxville, TN 37920;

[c] Department of Computer Science and Engineering, Texas A&M University,

College Station, TX, USA 77843

**Keywords:** mammography, visual search, eye-tracking, fractal analysis, diagnostic radiology error

PACS: 87.85.Ng 87.85.Tu

**Corresponding Author:**
Georgia D. Tourassi, PhD
Oak Ridge National Laboratory
1 Bethel Valley Road
PO Box 2008 MS6085
Oak Ridge, TN 37831-6085
Phone: (865) 576-4829
E-mail: tourassig@ornl.gov

## ABSTRACT

**Purpose**: The objective of this study is to assess the complexity of human visual search activity during mammographic screening using fractal analysis and to investigate its relationship with case and reader characteristics.

**Methods**: The study was performed for the task of mammographic screening with simultaneous viewing of four coordinated breast views as typically done in clinical practice. Eye-tracking data and diagnostic decisions collected for 100 mammographic cases (25 normal, 25 benign, 50 malignant) and 10 readers (three board certified radiologists and seven radiology residents), formed the corpus data for this study. The fractal dimension of the readers' visual scanning patterns was computed with the Minkowski–Bouligand box-counting method and used as a measure of gaze complexity. Individual factor and group-based interaction ANOVA analysis was performed to study the association between fractal dimension, case pathology, breast density, and reader experience level. The consistency of the observed trends depending on gaze data representation was also examined.

**Results:** Case pathology, breast density, reader experience level, and individual reader differences are all independent predictors of the visual scanning pattern complexity when screening for breast cancer. No higher order effects were found to be significant.

**Conclusions**: Fractal characterization of visual search behavior during mammographic screening is dependent on case properties and image reader characteristics.

## 1 INTRODUCTION

Breast cancer is the most frequently diagnosed form of cancer and the second leading cause of cancer-related deaths among women worldwide[1]. The mortality rate for this disease is largely dependent on early diagnosis through mammographic screening. With early detection, while the disease is localized, patients have a 98.5% relative survival rate versus 25% when the cancer is metastasized, a point at which the disease becomes incurable[2].

Previous studies showed the diagnostic interpretation of mammograms is susceptible to different types of human error resulting in missed diagnosis[3-7]. The topic of medical diagnostic error has received a lot of attention in recent years[8-10]. To this end, the medical research community has focused on the perceptual and cognitive processes related to decision making to better understand the causes of error. In radiology, misdiagnosis can be attributed to visual search and interpretation errors[11, 12].

For over half a century, a large number of studies have focused on the radiologists' visual scan pattern during medical image reading. Findings from these studies indicate prevalence of errors in two general categories: (1) how radiologists find what they are looking for (visual search); and (2) how radiologists interpret what they are looking at (image interpretation)[13-19]. A large body of eye-tracking research has also focused on gaining a better understanding of the relationship between visual search and diagnostic decision. These studies focus on analyzing radiologists' eye movements recorded during the diagnostic process[20-28].

In an early eye-tracking study of scanning strategies in mammography[14], Krupinski compared the eye-position data of six radiologists with variable experience levels (three experienced mammographers and three Radiology residents) to examine the influence of experience on scanning strategies. Krupinski found statistically significant differences between experienced and inexperienced image readers when comparing dwell time[14]. This study reported that less experienced image readers spend more time performing image search, creating a larger spatial coverage of the image (based on eye-position data)

when compared with more experienced image readers[14]. Kundel et al. investigated the concept of a global perceptual process during mammographic image reading[29]. They reported evidence of a global perceptual process in saccadic movements during the initial viewing of an image, and its importance in the identification of breast abnormalities[29]. They also found that more experienced radiologists develop a global perceptual process as a search strategy than their less experienced counterparts[29].

A more recent research study conducted by Voisin et al. showed the efficacy of eye-tracking in predicting diagnostic performance[30]. Voisin et al. conducted laboratory studies and applied machine learning techniques to predict error during the diagnostic characterization of mammographic lesions by combining features from radiologists' gaze behavior with textural image characteristics[30]. In a related study on breast cancer detection, Tourassi et al. investigated the relationship between radiologists' gaze, diagnostic decision, and image content of mammograms during mammographic screening[9]. Their results suggest that machine learning can be utilized to combine image content with the image reader's gaze characteristics to develop user-dependent models for predicting medical error in breast cancer lesion detection and characterization[9].

In an earlier study of eye-movements, Engbert and Kliegl[31] investigated the function of microsaccades during visual fixation on stationary targets. Using a standard deviation analysis, the authors showed that microsaccades are triggered dynamically. This triggering mechanism allows for prediction of individual microsaccade rate using fractal dimension of trajectories. A more recent study by Stephen and Anastas[32] examined fractal fluctuations in speed during visual search. They analyzed the fractality of angular changes in gaze and reported finding fractal fluctuations during visual search. They also reported correlations between cognitive performance (measured as reaction time during visual search) and fractality of gaze.

Although a number of investigators have examined radiologists' visual scanning patterns for screening mammograms, the discovered patterns are typically summarized with respect to features such as total time examining a case, time to initially hit true lesions, total dwell time, number of hits, etc. While

105 informative, these features fail to capture the gaze path trajectory and therefore they cannot fully capture the complexity of the visual search process. In addition, earlier studies were based on single view

Table 1. Specifications of the 100 mammographic cases used in the study.

| Ground Truth | Patient Age | Breast Density | Mass Subtlety | Total Abnormalities | No. of Cases |
|---|---|---|---|---|---|
| **Normal** | Range: 36 – 68 (56.2 ± 10.6) | Range: 1 – 4 (Median: 2) | N/A | N/A | 25 |
| **Benign** | Range: 34 – 82 (56.9 ± 13.4) | Range: 1 – 3 (Median: 2) | Range: 3 – 5 (Median: 5) | Range: 1 – 3 (Median: 1) | 25 |
| **Malignant** | Range: 37 – 83 (64.3 ± 12.4) | Range: 1 – 4 (Median: 2) | Range: 1 – 5 (Median: 5) | Range: 1 – 3 (Median: 1) | 50 |

mammograms, which is not consistent with clinical practice. Mammographic screening entails simultaneous viewing of 4 coordinated breast views. The purpose of our study is to address the limitations of the earlier investigations and attempt to characterize the complexity of the radiologists' visual search
110 activity when viewing 4-view mammographic cases as a function of three factors: (i) breast parenchyma density, (ii) case pathology, and (iii) radiologists' experience level. Our study however, focused on mass detection, which is known to be associated with higher detection error than microcalcifications[3, 4, 6, 13].

In a previous study[33], we presented results on the efficacy of fractal analysis as a measure of visual search complexity based on a default viewing configuration for each reader on all mammographic cases
115 as a proof of concept. Our work here represents a more rigorous analysis of this approach. We examined the fractal signature of visual search patterns for all valid display configurations to eliminate the possibility of reporting trends resulting from systematic bias. In addition, this work presents deeper analysis to evaluate the reader-specific robustness of the observed trends based on each image reader's dominate viewing configuration on a per case basis.

120 The contributions of this paper include: (i) fractal analysis of scanpath to assess of complexity of visual search, and (ii) generation and comparison of differences between visual search complexity profile

of individual readers. To the best of our knowledge, there are no previous studies, which report the application of fractal dimension as a metric for assessing complexity of scanpath during visual search in mammography as highlighted in this paper.

## 2 MATERIALS AND METHODS

### 2.A. Image Database

To perform this study, 100 screen-film mammograms were selected from a corpus of mammographic cases digitized with a high resolution LUMISYS scanner (50μm per pixel, 12 bit) from the University of South Florida's Digital Database for Screening Mammography (DDSM)[34]. Each DDSM case contains 4 images, the craniocaudal (CC) and mediolateral oblique (MLO) views of both the left and the right breasts, associated ground truth established via biopsy, additional imaging, or 2-year follow-up, radiologist's assessment using the BI-RADS™ lexicon[35], and patient age.

The selected set included clinically actionable cases covering a broad range of mass margin and shape characteristics. The cases were selected at random. The mass cases were selected with restriction criteria based on mass annotation size. Specifically, we selected DDSM mass cases for which the bounding box of the provided mass annotations was less than 512 x 512 pixels. This effectively translates into masses less than 15mm in radius. Please note that the DDSM mass annotations tend to be generous without tracing the actual mass outline. Therefore, the 512 x 512 pixel size is an extreme upper limit of the masses included in our experiments. Fifty cases included biopsy-proven malignant masses, 25 cases included biopsy-proven benign masses, and the remaining 25 cases were normal as determined during a

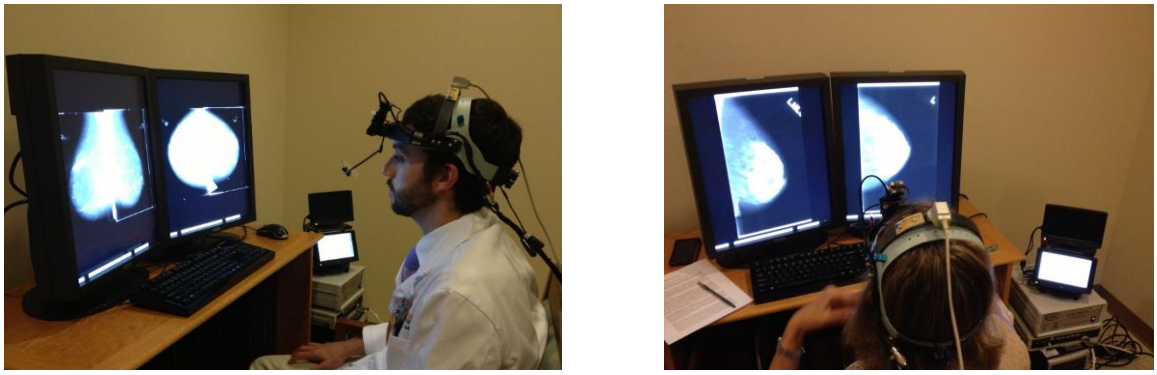
Table 2. Summary of characteristics of study participants

| Reader Type | Experience Level | No. of Participants |
|---|---|---|
| Radiologist | ≥ 9 yrs of practice | 3 |
| Advanced Resident | > 2 mammo rotations | 4 |
| New Resident | 1 – 2 mammo rotations | 3 |

2-year cancer-free follow-up patient evaluation. Mammograms with masses deemed as "benign-without-callback" were excluded. The overwhelming majority of the mass cases (72 out of 75) did not include any microcalcifications. Mass conspicuity was assessed according to the subtlety rating provided in the DDSM truth files. These ratings ranged from 1 (suggesting a subtle lesion) to 5 (suggesting an obvious lesion). A complete list of the DDSM cases used in this study is provided in the Appendix. Table 1 provides details on the selected cases, including information on the patient's age and breast parenchymal density. The parenchymal density is also provided in the DDSM truth files and it ranged between 1 (fatty) to 4 (dense), according to the BI-RADS™ lexicon[35].

### 2.B. Data Collection Protocol

Ten readers of variable experience levels from an academic institution were recruited to conduct blind review of the selected mammograms. Each reader was asked to report the location of any suspicious mass and provide a corresponding BI-RADS rating as typically done in clinical practice. Of the ten readers, three were experienced MQSA-certified radiologists each with at least nine years of dedicated mammographic experience, four radiology residents with at least three mammography rotations, and three radiology residents with 1 or 2 mammography rotations (see Table 2). Institutional review board approval was obtained prior to the study. Human subject recruitment and data collection was done according to a protocol approved by the Oak Ridge Site-Wide Internal Review Board. All participants signed an informed consent form.



(a) (b)

Figure. 1. Reader outfitted with eye-tracking apparatus reviewing 4-view mammographic case. (a) Dual Display showing default arrangement. (b) Dual display showing LMLO and LCC views on left and right displays respectively.

A customized graphical user interface (GUI) was developed in-house for study participants to view each mammographic case and record their findings. Two medical grade monitors were used (dual-head 5MP mammo-grade Totoku LCD monitors calibrated to the DICOM display standard). The four mammographic views (LCC, RCC, LMLO, RMLO) were initially displayed at low resolution (two views per monitor) to fit the screen. To assess breast symmetry, the users could select the MLO views to be displayed on the left monitor and the CC views to be displayed on the right monitor (Fig. 1a). The readers were also able to select and view a single breast at full spatial resolution with the MLO view displayed on the left monitor and the CC view displayed on the right monitor (e.g., Fig. 1b). Table 3 enumerates all possible "hanging protocols" implemented in the GUI. Please note that based on the allowable protocols RMLO could never appear on the right monitor while LCC could never appear on the left monitor. In addition, the GUI provided the functionality of zooming in/out, panning, and magnifying glass for detailed reading of each mammographic view. During the reading sessions, each reader was outfitted with an H6 head-mounted eye-tracker, with a 60 Hz sampling rate, and eye-head integration from Applied Science Laboratories (ASL, Bedford, Massachusetts, USA). The eye-tracker recorded each reader's eye position data to within 0.5º of accuracy.

Readers were instructed to take as much time as needed to view each case until they were satisfied with the viewing phase. Readers were also informed about the presence of both normal and abnormal

Table 3**.** Enumeration of dual display viewing arrangements and corresponding images on each monitor

| **Dual Display Viewing Arrangements** | **Left Monitor** | **Right Monitor** |
|---|---|---|
| 1: Same mammographic view (CC) | RCC | LCC |
| 2: Same mammographic view (MLO) | RMLO | LMLO |
| 3: Same breast viewing (Right) | RMLO | RCC |
| 4: Same breast viewing (Left) | LMLO | LCC |
| 5: Four-view (default) | RMLO & LMLO | RCC & LCC |

cases but no information was provided to them regarding the expected prevalence. Once the reader was prepared to give a diagnostic assessment of the case, the eye-tracking recording process was halted pending complete reporting of case specific findings, and the reader was ready to proceed with viewing the next case. The reader's reporting task was to mark and rate any suspicious findings. Each mark was classified and rated for likelihood of malignancy on a BI-RADS-based scale, which consists of five levels (2, 3, 4A, 4B, 4C, and 5) of increasing probability of malignancy. Cases with no markings were assigned a BI-RADS rating of 1. After completion of case reporting, the reader was instructed to proceed with the next case. Prior to the every reading session, each reader was carefully calibrated using the 9-point calibration protocol provided by ASL and trained on five training cases selected from the DDSM database. The set of training cases were different from the 100 test cases used in the study.

The cases were presented in a randomized order. A different randomization scheme was used for each reader. Readers were also permitted to complete the study in multiple sessions based on preference and scheduling conflicts. For example, of the ten readers, two completed case readings for the study in one day (over two sessions), four completed case readings for the study in two days (over at most three sessions), and one reader completed case readings in three days (over 4 sessions).

### 2.C. Data Processing

As described in the previous section, gaze data for each reader and each case were collected from four mammographic views spread across two monitors. Raw gaze data was preprocessed using the EyeNAL analysis program from Applied Science Laboratory, which converts raw gaze data to a time-ordered sequence of fixations $f_1$, $f_2$,…,. $f_n$., along with other measures associated with fixation (such as fixation duration and inter-fixation degree). These fixations represent a grouping of at least three temporally sequenced raw gaze-position points within 0.5° of visual angle of each other, and a minimum threshold of 100ms total gaze time.

The scanpath, derived by connecting time-ordered fixations or gaze points while viewing each case, resulted in a dense gaze scanpath. To measure the complexity of this graph we used the scalar quantity fractal dimension (***D***). Fractal geometry is superior to Euclidian geometry for describing complex, rough, irregular and often branching objects, which occur naturally[36]. The non-integer, fractal dimension, which measures the fractality of an object is the fundamental metric used in fractal geometry. Fractal dimension (D) has been used in various areas of science, predominantly Biology, to characterize the complexity of shapes in animal and plant morphology[37, 38]. Fractal analysis has also been applied in studying complexity in search behavior patterns in marine predators[39], honey bees[40], and other animals[41].

The gaze scanpath can be treated as a fractal pattern. Its fractal dimension is a non-integer ***D*** with the range: ($n - 1 < D \leq n$) where n=2 is the pattern dimensionality. Using the Minkowski–Bouligand box-counting method[42], we estimated ***D*** for each gaze scanpath graph derived from each case examined by each reader. Suppose $\boldsymbol{N(\varepsilon)}$ is the number of boxes of length ε required to cover the gaze scanpath ***G***, we define $\boldsymbol{D_{box}}$ for the two-dimensional graph as:

$$D_{box}(G) := \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log(1/\varepsilon)} \quad (1)$$

## 2.D. Image Representation and Visual Search

The first step in preprocessing was to combine data from the ASL eye-tracking apparatus, user device interactions including mouse interactions, and alternating between views into a single unified time and coordinate space. During user interaction, the coordinates of the eye-tracker was captured in physical units (e.g. inches), while user interactions were recorded in computer display coordinates (pixels). The resulting heterogeneous coordinates were first translated into a unified coordinate to perform eye tracking analyses on multiple displays. To achieve this, we translated both coordinates into image pixel coordinates. Since the display software stored time-synchronized information about user interactions, including zooming in and out, and switching view, we mapped data from eye tracking apparatus to the underlying image pixel coordinate. Through this process, eye gaze data captured during experiments, locations of user interactions such as mouse clicks and drags as image readers provided BI-RADS ratings, were translated into a unified image pixel coordinate space and time.

During the reading session, readers typically jump from one of the five possible dual display viewing arrangements (see Table 3) to another resulting in a unique non-homogeneous two-dimensional image coordinate space of eye position data for each display view arrangement. To perform fractal analysis of gaze patterns, raw eye position data from each of the unique coordinate spaces was combined to create a single two-dimensional coordinate space, representing eye-position data for each individual case. Table 4 enumerates six possible configurations for data representation (i.e., configurations $C_i$, i=1, 2,…,6) for aggregating gaze data into a single coordinate space based on the allowable hanging protocols. The default data representation is the one that corresponds on the default hanging protocol applied at the beginning of each case as illustrated in Figure 1a (i.e. configuration $C_6$ from Table 4, RMLO|LMLO|RCC|LCC).

Table 4**.** Possible configurations for a combined two-dimensional data representation.

| | **Left Monitor** | | **Right Monitor** | |
|---|---|---|---|---|
| | **Left Image** | **Right Image** | **Left Image** | **Right Image** |
| $C_1$ | Right (RMLO) | Right (RCC) | Left (LMLO) | Left (LCC) |
| $C_2$ | Right (RMLO) | Right (RCC) | Left (LCC) | Left (LMLO) |
| $C_3$ | Right (RCC) | Right (RMLO) | Left (LMLO) | Left (LCC) |
| $C_4$ | Left (LMLO) | Right (RMLO) | Right (RCC) | Left (LCC) |
| $C_5$ | Right (RMLO) | Left (LMLO) | Left (LCC) | Right (RCC) |
| $C_6$ | Right (RMLO) | Left (LMLO) | Right (RCC) | Left (LCC) |

We converted raw eye position data for the duration of each case in two steps. First, we mapped gaze position onto a mammographic image-dependent pixel coordinate space to handle zoom, image translation, and other artifacts from eye tracking. Subsequently, each mammographic image, along with respective eye position data were mapped onto a unified pixel coordinate space through a simple translation and scaling (see EQ 2.).

$$\begin{bmatrix} \acute{x} \\ \acute{y} \\ 1 \end{bmatrix} = \begin{bmatrix} A\cos\theta & -A\sin\theta & d_{xi} \\ A\sin\theta & A\cos\theta & d_{yi} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2)$$

where $A$ represents a scaling factor, $\theta$ represents an angle of rotation (set to zero for our purposes), and $d_{xi}$ and $d_{yi}$ represent translation parameters for the $i^{th}$ mammographic image.

Initial analysis was performed on the data representation corresponding to the default image arrangement (see configuration $C_6$ in Table 4). Further, we investigated the effects, if any, of using alternative configurations for data representation (see $C_1 – C_5$ in Table 4) on the computed fractal dimension and if any discovered effects alter our initial findings.

## 2.E. Observer Performance Measurement

In this study, we compare performance of radiologists of varied experience levels in the abnormality detection task with the various images. First, we mapped the diagnostic decision for each

case to a linear scale based on the BI-RADS rating provided. We designated cases without markings (i.e. no scores were given) as 0; BI-RADS ratings {2 and 3} 1 and 2 respectively; and BI-RADS ratings {4A, 4B, 4C, and 5} as 3,4,5, and 6 respectively. Using this linear rating scale, the receiver operating characteristic curve (ROC) analysis was performed using a web-based analysis tool[43] and report the area under the ROC curve (AUC) for each image reader.

In addition, to determine mass detection accuracy, we compared the BI-RADs ratings provided by each reader with the ground truth. We grouped benign and malignant cases under a single class label: mass present (M), and normal cases under a second class label: mass absent / normal (N). We report the average diagnostic accuracy using this two-class grouping (mass-present vs. mass-absent) for each image reader.

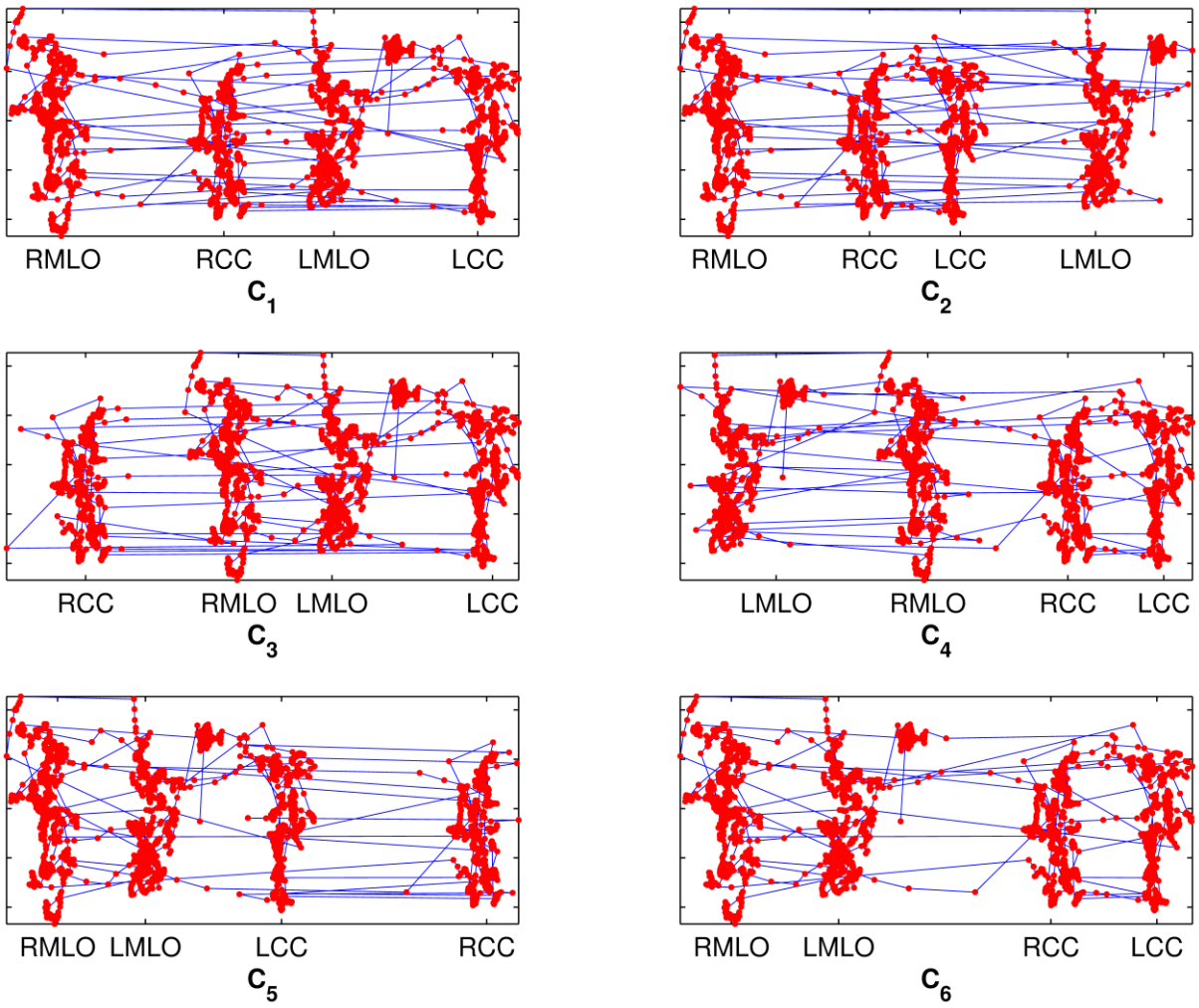



Figure 2. Gaze data collected for a single reader synthesized in the 6 possible configurations for data representation.

270

Table 5. Mass detection performance (mass-present vs. mass-absent) for new residents (NR), advanced resident (AR), and expert (E) radiologists.

| | NR1 | NR2 | NR3 | AR1 | AR2 | AR3 | AR4 | E1 | E2 | E3 |
|---|---|---|---|---|---|---|---|---|---|---|
| **True Positive** | 59 | 71 | 62 | 49 | 47 | 49 | 38 | 75 | 72 | 72 |
| **True Negative** | 12 | 3 | 5 | 18 | 14 | 17 | 17 | 5 | 10 | 9 |
| **False Positive** | 13 | 22 | 20 | 7 | 11 | 8 | 8 | 20 | 15 | 16 |
| **False Negative** | 16 | 4 | 13 | 26 | 28 | 26 | 37 | 0 | 3 | 3 |
| **Sensitivity** | 0.79 | 0.95 | 0.83 | 0.65 | 0.63 | 0.65 | 0.51 | 1.00 | 0.96 | 0.96 |
| **Specificity** | 0.48 | 0.12 | 0.20 | 0.72 | 0.56 | 0.68 | 0.68 | 0.20 | 0.40 | 0.36 |
| **Accuracy** | 0.71 | 0.74 | 0.67 | 0.67 | 0.61 | 0.66 | 0.55 | 0.80 | 0.82 | 0.81 |
| **AUC ROC** | 0.77 | 0.77 | 0.74 | 0.76 | 0.71 | 0.68 | 0.71 | 0.87 | 0.84 | 0.88 |

# 3 RESULTS

## 3.A. Radiologists' Diagnostic Performance

We grouped each of the 10 participating readers into one of three experience levels: new trainee resident (NR), advanced trainee resident (AR), and expert radiologist (E) as illustrated in Table 2. We mapped the diagnostic decision for each case to one of the three case pathologies (normal, benign, malignant) based on the BI-RADS rating provided. We designated cases without markings (i.e. no scores were given) as normal (N); we grouped BI-RADS ratings {2 and 3} as benign (B); and we grouped BI-RADS ratings {4A, 4B, 4C, and 5} as malignant (M). We formed three breast parenchyma density groupings by combining heterogeneous and dense cases in the same density grouping (due to the small sample size of density 4).

Each reader was asked to report the location of any suspicious mass and provide a corresponding BI-RADS rating as typically done in clinical practice. The diagnosis was deemed correct if the BI-RADS rating matched the ground-truth pathology of the case (as outlined in the previous paragraph), and the

location of the markings provided were within the DDSM provided mass annotation (where applicable). In cases where more than one lesion was present, the case was deemed correctly diagnosed if the correct location and rating was provided for at least one of the lesions.

To determine mass detection accuracy, we compared the BI-RADs ratings provided by each reader with the ground truth. We grouped benign and malignant cases under a single class label: mass present (M), and normal cases under a second class label: mass absent / normal (N). We report the average diagnostic performance using this two-class grouping (mass-present vs. mass-absent) for each individual radiologist in Table 5. From Table 5, we deduce the average accuracy by experience level: 70.7%±3.5% (new residents), 62.25%±5.5% (advanced residents) and 81%±1.0% (experts). The accuracy of the expert radiologists was significantly higher than that of the advanced residents, $t(5) = 5.7$, $p = 0.002$, and the new residents, $t(4) = 4.9$, $p = 0.008$. No significant difference accuracy was observed between new residents and advanced residents, $t(5) = 2.29$, $p = 0.07$. Readers appeared to execute the clinical task by operating with very different decision criteria in terms of emphasizing sensitivity vs. specificity.

To compare abnormality detection performance, we applied the linear rating scale described in Section 2.E., to compute the area under the ROC curve (AUC) for each image reader. A summary of the results is provided in Table 5. The average group level AUC followed the same trend as observed in the mass detection accuracy scores. The average AUC for experienced radiologist group (0.863±0.021), was higher, $t(4) = 6.61$, $p = 0.003$, than the new resident group (0.76±0.017), and higher, $t(5) = 6.73$, $p = 0.001$, than the advanced resident group (0.715±0.033). No significant difference was observed between the two resident groups, $t(5) = 2.11$, $p = 0.09$.

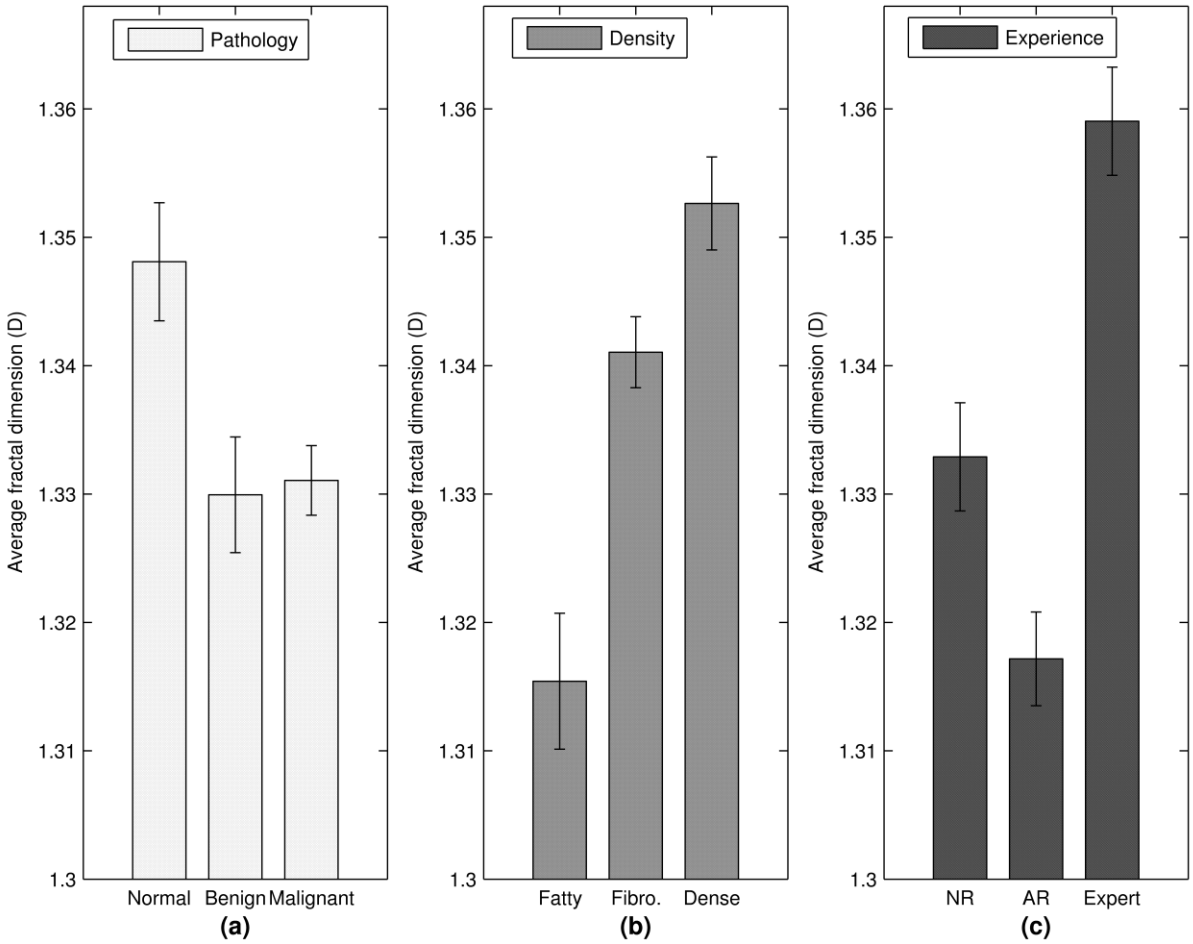



Figure 3. Averaged complexity of visual search across case and reader properties: (a) case pathology (normal, benign, and malignant); (b) breast density (fatty, fibroglandular, and heterogeneous/dense); and (c) image reader experience level: new Radiology residents (NR); advanced Radiology residents (AR), and expert radiologists (E).

### 3.B. Fractal Dimension of Radiologists' Gaze Scanpath

The fractal dimension of the readers' gaze scanpath ranged between 1.08 and 1.51. In Figure 3, we present the average fractal dimension across all cases grouped by case specific properties: case pathology (normal, benign, and malignant)), breast density (fatty, fibroglandular, and heterogeneous/dense), and readers' experience level (new Radiology resident, advanced Radiology resident, and expert radiologist).

***Effect of Case Pathology on Complexity of Visual Search***. The average complexity of gaze for normal cases (1.350 ± 0.005) was significantly higher, $t(498) = 3.37$, $p < 0.01$, than the average complexity for

mass-present cases, which contain a benign mass (1.330 ± 0.005), and similarly higher, $t(748) = 5.05$, $p < .01$) than the average complexity for mass-present cases, which contain a malignant mass (1.331 ± 0.003). However, there was no significant difference, $t(748) = 1.17$, $p = 0.98$), in the average complexity of gaze between malignant and benign cases.

***Effect of Mammographic Density on Complexity of Visual Search***. In Figure 3b, we observe that the complexity of gaze increases monotonically with mammographic density. The average complexity of gaze (1.315 ± 0.006) for low-density mammographic cases is significantly lower, $t(718) = 5.67$, $p < 0.001$), compared with the average complexity (1.340 ± 0.003) for medium-density mammographic cases. The average complexity of gaze for low-density images was also significantly lower , $t(528) = 7.37$, $p << 0.001$), compared with the average complexity (1.353 ± 0.004) for high-density (heterogeneous/dense) cases. The average complexity of gaze for medium-density mammographic cases was also significantly lower, $t(528) = 5.67$, $p = 0.02$, than the gaze complexity for high-density mammographic cases.

***Effect of Readers' Experience Level on Complexity of Visual Search***. Figure 3c illustrates the averaged complexity of gaze for image readers grouped by experience level. We observe that the average complexity of gaze for experienced radiologists (1.360 ± 0.004) is significantly higher, $t(598) = 4.29$, $p < 0.001$, than the average complexity for new Radiology residents (1.330 ± 0.004), and significantly higher, $t(698) = 7.54$, $p << 0.001$, than the average complexity for advanced Radiology residents (1.320 ± 0.003). The average gaze complexity of advanced Radiology residents was significantly lower, $t(528) = 7.37$, $p = 0.01$, than that of new Radiology residents.

Table 6. Multi-factor ANOVA test results for possible configurations for data representation.

| Source | Configuration | | | | | |
|---|---|---|---|---|---|---|
| | $C_1$ | $C_2$ | $C_3$ | $C_4$ | $C_5$ | $C_6$ |
| | **p > F** | **p > F** | **p > F** | **p > F** | **p > F** | **p > F** |
| Pathology | << 0.001 | << 0.001 | << 0.001 | << 0.001 | << 0.001 | << 0.001 |
| Density | << 0.001 | << 0.001 | << 0.001 | << 0.001 | << 0.001 | << 0.001 |
| Experience | << 0.001 | << 0.001 | << 0.001 | << 0.001 | << 0.001 | << 0.001 |
| Individual | << 0.001 | << 0.001 | << 0.001 | << 0.001 | << 0.001 | << 0.001 |
| Pathology – Density | 0.92 | 0.93 | 0.91 | 0.86 | 0.92 | 0.88 |
| Pathology – Experience | 0.31 | 0.32 | 0.21 | 0.24 | 0.29 | 0.32 |
| Pathology – Individual | 0.16 | 0.1 | 0.14 | 0.11 | 0.07 | 0.11 |
| Density – Experience | 0.62 | 0.83 | 0.72 | 0.78 | 0.8 | 0.85 |
| Density – Individual | 0.06 | 0.11 | 0.11 | 0.06 | 0.1 | 0.03 |
| Pathology – Density – Experience | 0.58 | 0.91 | 0.87 | 0.88 | 0.91 | 0.48 |
| Pathology – Density – Individual | 0.53 | 0.85 | 0.77 | 0.8 | 0.85 | 0.32 |

### 3.C. Analysis of Variations in Visual Search Complexity

The gaze scanpaths generated during mammographic screening varied in complexity with the characteristics of each case (pathology and density) and with individual radiologists (as observed in Figure 3). Therefore, we performed ANOVA on the fractal dimensions for each case to determine if there is a dependency with case pathology, breast density, or reader experience level. To analyze the interaction between gaze complexity, case pathology, case density, and reader experience level, we applied a four-factor fixed-effects ANOVA with three levels for case pathology (normal, benign, and malignant), three levels for breast parenchyma density (fatty, fibroglandular, and heterogeneous/dense), and three experience levels (new trainee, advanced trainee, and expert), across 10 individual readers. In Table 6, we report ANOVA test results using fractal dimensions estimated for the six image configurations illustrated in Table 4. ANOVA showed that all four factors are independent predictors of a radiologists' visual search complexity. The overall results were consistent across all six configurations for data

Table 7 Pairwise comparisons on groups of case pathology, breast density, and reader experience level

| Pair 1 | Pair 2 | p–value |
|---|---|---|
| Pathology – Normal | Pathology – Benign | 0.01 |
| Pathology – Normal | Pathology – Malignant | 0.001 |
| Pathology – Benign | Pathology – Malignant | 0.98 |
| Density – Fatty | Density – Fibroglandular | 0.005 |
| Density – Fatty | Density – Heterogeneous/Dense | 0.003 |
| Density – Fibroglandular | Density – Heterogeneous/Dense | 0.004 |
| Experience – New Trainee | Experience – Advanced Trainee | 0.01 |
| Experience – New Trainee | Experience – Expert | << 0.001 |
| Experience – Advanced Trainee | Experience – Expert | << 0.001 |

representation, with the exception of one higher order effect (Density – Individual on configuration $C_6$), which was found to be significant ($F(14,910) = 2.02$, $p = 0.03$).

Overall, ANOVA results show that the pathology and density of a mammographic case both have a significant effect, $F(2, 910) = 18.77$, $p < 0.001$ and $F(2, 910) = 33.57$, $p << 0.001$ respectively, on visual search complexity as calculated using fractal dimension. The ANOVA tests also show that individual factors (individual differences and level of experience) both have a significant effect, $F(7, 910) = 47.82$, $p < 0.001$, and $F(2, 910) = 43.16$, $p << 0.001$ respectively, on fractal dimension. These findings indicate that the trends observed in Figure 3 (and highlighted in section 3.B.) are significant.

Since ANOVA results did not depend on the configuration used for data representation, we used a case-dependent data representation to compute visual search complexity for each case. This approach computed the visual search complexity from raw gaze data based on the predominant display arrangement used by the reader for each case. We applied five-factor fixed-effects ANOVA on the case-dependent visual search complexity by including readers' diagnostic interpretation as the fifth factor along with pathology, density, experience, and individual differences (as described in section 3.B). The results of ANOVA tests were consistent with our previous findings. However, this analysis showed that the reader's diagnostic decision is an independent predictor of visual search complexity $F(2,923) = 6.62$, $p < 0.01$.

Table 8. Pairwise comparisons of individual readers (new resident (NR), advanced resident (AR), and expert (E)).

| | **NR1** | **NR2** | **NR3** | **AR1** | **AR2** | **AR3** | **AR4** | **E1** | **E2** |
|---|---|---|---|---|---|---|---|---|---|
| **NR2** | < 1e-3 | | | | | | | | |
| **NR3** | 0.99 | < 1e-3 | | | | | | | |
| **AR1** | < 1e-3 | 0.77 | 0.01 | | | | | | |
| **AR2** | 0.57 | < 1e-3 | 0.99 | 0.27 | | | | | |
| **AR3** | < 1e-3 | 0.32 | < 1e-3 | 0.001 | < 1e-3 | | | | |
| **AR4** | 0.89 | < 1e-3 | 1.0 | 0.077 | 1.0 | < 1e-3 | | | |
| **E1** | 0.005 | 0.3 | 0.12 | 1.0 | 0.75 | < 1e-3 | 0.39 | | |
| **E2** | 0.9 | < 1e-3 | 0.34 | < 1e-3 | 0.02 | < 1e-3 | 0.097 | < 1e-3 | |
| **E3** | 0.002 | < 1e-3 | < 1e-3 | < 1e-3 | < 1e-3 | < 1e-3 | < 1e-3 | < 1e-3 | 0.22 |

Post-ANOVA t-tests with Bonferroni p-value adjustment were also performed and reported in Table 7. The complexity of the readers' visual search was significantly different between normal cases and mass-present cases. However, the malignancy status of a mass did not affect the complexity of the readers' visual search. Further, visual search complexity was found to be significantly different between mammograms of fatty breasts and mammograms of fibroglandular and heterogeneous/dense breasts. However, there was no significant difference in visual search complexity between mammograms of fibroglandular breasts and heterogeneous/dense breasts. We also observed that visual search complexity was significantly different between all three experience groups: new Radiology residents, advanced Radiology residents and expert radiologists.

Finally, a paired-sample t-tests was conducted to compare the pairwise differences in complexity of gaze scanpaths among the 10 readers (Table 8). Several significant pairwise differences were found suggesting that there is substantial inter-reader variability, often among readers of similar experience level.

## 4 DISCUSSION

This study investigated the efficacy of visual gaze complexity for characterizing the search behavior of radiologists when viewing mammograms for breast cancer screening. For this study fractal dimension was used as the metric for quantifying the complexity of the visual search patterns. Using a relatively large number of cases, comprised of varied pathology and breast parenchyma density, and image readers with varied levels of experience and expertise, the findings presented in this study suggest the following trends:

(1) The characteristics of a mammographic case (pathology and breast parenchyma density) are independent factors in predicting complexity of visual search behavior.

(2) The characteristics of the image reader (individual differences and level of experience) are independent factors in predicting complexity of visual search behavior.

(3) The pathology and breast parenchyma density of a mammographic case, experience level of the image reader, and the resulting diagnostic decision are combined predictors of visual search complexity during mammographic screening.

(4) Visual search complexity is significantly different between normal and mass-present cases.

(5) The visual search complexity increases monotonically with increasing breast parenchyma density. Effectively, low-density mammographic images correspond to lower visual search complexity, while medium-density images correspond to a higher visual search complexity, and high-density images correspond to the highest visual search complexity. This finding is consistent with results obtained by Al Mousa et al.[44], who reported significant increases in visual search parameters when comparing low- and high-density mammograms.

(6) On average, the visual search complexity of Radiology residents (both the new and the advance trainee groups) are significantly lower than the average complexity of experienced radiologists.

(7) There are notable differences in visual search complexity between individual radiologists.

This study is novel in its replication of the dual monitor viewing and decision tasks that are characteristic of screening mammography in practice. It presents a single quantity, fractal dimension, capturing the complexity of visual search behavior during the mammographic screening process. This metric can be further investigated as a feature to develop more accurate models for predicting individualized radiologist error risks for a specific case in review. These findings also present future research opportunities in personalized decision support and training support technology in Radiology.

Despite the replication of dual monitor viewing and decision tasks, which are characteristic of screening mammography in practice, there are notable limitations with this study. While fractal dimension successfully characterizes spatial complexity of visual search, it does not incorporate any temporal information which, intuitively, contain information relevant to readers' visual search behavior and diagnostic performance as noted in[44-46]. We are currently working on developing novel strategies to capture such information. In addition, our study focused specifically on the detection of mammographic masses. It is important to investigate the same issue for other mammographic lesions as well. Lastly, our study utilized a popular but fairly old dataset of digitized mammograms.

By leveraging a publicly available dataset that has been extensively used by the research community, other researchers will be able to reproduce our experimental design and perform comparative studies with of new visual search analysis algorithms based on the same list of DDSM cases we used. Still, a separate study is needed to confirm how our findings would translate in digital mammography. A prior study suggested significant differences in visual scan behavior between screen-film and digital mammograms[11]. However, that earlier study was based on two-view mammograms (single breast viewing) without any ability for zooming. Furthermore, the differences observed in that study involved traditional metrics such as time to first hit and total dwell time. Our study implemented a clinically realistic viewing scenario and a more spatially comprehensive metric of visual search. Furthermore, by providing the full list of the publicly available cases we used we enable other researchers to perform comparative studies.

## 5 ACKNOWLEDGEMENT

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

## 6 REFERENCES


[1] A.B. Ryerson, C.R. Eheman, S.F. Altekruse, J.W. Ward, A. Jemal, R.L. Sherman, S.J. Henley, D. Holtzman, A. Lake, A.-M. Noone, R.N. Anderson, J. Ma, K.N. Ly, K.A. Cronin, L. Penberthy, B.A. Kohler, "Annual Report to the Nation on the Status of Cancer, 1975-2012, featuring the increasing incidence of liver cancer," Cancer **122**, 1312-1337 (2016).

[2] R.A. Smith, V. Cokkinides, A.C. von Eschenbach, B. Levin, C. Cohen, C.D. Runowicz, S. Sener, D. Saslow, H.J. Eyre, "American Cancer Society Guidelines for the Early Detection of Cancer," CA: A Cancer Journal for Clinicians **52**, 8-22 (2002).

[3] R.E. Bird, T.W. Wallace, B.C. Yankaskas, "Analysis of cancers missed at screening mammography," Radiology **184**, 613-617 (1992).

[4] L. Berlin, "Malpractice issues in radiology. Perceptual errors," American Journal of Roentgenology **167**, 587-590 (1996).

[5] L. Berlin, "Malpractice issues in radiology. Errors in judgment," AJR-American Journal of Roentgenology **166**, 1259-1261 (1996).

[6] B.C. Yankaskas, M.J. Schell, R.E. Bird, D.A. Desrochers, "Reassessment of Breast Cancers Missed During Routine Screening Mammography," American Journal of Roentgenology **177**, 535-541 (2001).

[7] K. Ganesan, U.R. Acharya, C.K. Chua, L.C. Min, K.T. Abraham, K.-H. Ng, "Computer-Aided Breast Cancer Detection Using Mammograms: A Review," IEEE Reviews in Biomedical Engineering **6**, 77-98 (2013).

[8] C.S. Lee, P.G. Nagy, S.J. Weaver, D.E. Newman-Toker, "Cognitive and System Factors Contributing to Diagnostic Errors in Radiology," American Journal of Roentgenology **201**, 611-617 (2013).

[9] G. Tourassi, S. Voisin, V. Paquit, E. Krupinski, "Investigating the link between radiologists' gaze, diagnostic decision, and image content," Journal of the American Medical Informatics Association **20**, 1067-1075 (2013).

[10] M.L. Graber, S. Kissam, V.L. Payne, A.N.D. Meyer, A. Sorensen, N. Lenfestey, E. Tant, K. Henriksen, K. LaBresh, H. Singh, "Cognitive interventions to reduce diagnostic error: a narrative review," BMJ Quality & Safety2012).

[11] R.L. Birdwell, D.M. Ikeda, K.F. O'Shaughnessy, E.A. Sickles, "Mammographic Characteristics of 115 Missed Cancers Later Detected with Screening Mammography and the Potential Utility of Computer-aided Detection," Radiology **219**, 192-202 (2001).

[12] E.A. Krupinski, "Current perspectives in medical image perception," Attention, Perception, & Psychophysics **72**, 1205-1217 (2010).

[13] M.D. Mugglestone, A.G. Gale, H.C. Cowley, A.R.M. Wilson, "Defining the perceptual processes involved with mammographic diagnostic errors," in *Medical Imaging*, *Vol. 2712* (International Society for Optics and Photonics, 1996), pp. 71-77.

[14] E.A. Krupinski, "Influence of experience on scanning strategies in mammography," in *Medical Imaging 1996*, *Vol. 2712* (International Society for Optics and Photonics, 1996), pp. 95 -101.

[15] C. Mello-Thoms, S. Dunn, C.F. Nodine, H.L. Kundel, S.P. Weinstein, "The Perception of Breast Cancer: What Differentiates Missed from Reported Cancers in Mammography?," Academic Radiology **9**, 1004-1012 (2002).

[16] C. Mello-Thoms, "How Does the Perception of a Lesion Influence Visual Search Strategy in Mammogram Reading?," Academic Radiology **13**, 275-288 (2006).

[17] C. Mello-Thoms, "The problem of image interpretation in mammography: effects of lesion conspicuity on the visual search strategy of radiologists," The British Journal of Radiology **79**, S111-S116 (2006).

[18] C. Mello-Thoms, "Visual search characteristics in mammogram reading: SFM vs. FFDM," in *SPIE Medical Imaging* (International Society for Optics and Photonics, 2010), pp. 76270D-76270D-76278.

[19] S.L. Rose, A.L. Tidwell, M.F. Ice, A.S. Nordmann, R. Sexton Jr, R. Song, "A Reader Study Comparing Prospective Tomosynthesis Interpretations with Retrospective Readings of the Corresponding FFDM Examinations," Academic Radiology **21**, 1204-1210 (2014).

[20] H.L. Kundel, D.J. Wright, "The Influence of Prior Knowledge on Visual Search Strategies During the Viewing of Chest Radioqraphs," Radiology **93**, 315-320 (1969).

[21] H.L. Kundel, J. Paul S. La Follette, "Visual Search Patterns and Experience with Radiological Images," Radiology **103**, 523-528 (1972).

[22] H.L. Kundel, C.F. Nodine, D. Carmody, "Visual Scanning, Pattern Recognition and Decision-making in Pulmonary Nodule Detection," Investigative Radiology **13**, 175-181 (1978).

[23] H.L. KUNDEL, C.F. NODINE, E.A. KRUPINSKI, "Searching for Lung Nodules: Visual Dwell Indicates Locations of False-Positive and False-Negative Decisions," Investigative Radiology **24**, 472-478 (1989).

[24] D.J. Manning, S.C. Ethell, T. Donovan, "Detection or decision errors? Missed lung cancer from the posteroanterior chest radiograph," The British Journal of Radiology **77**, 231-235 (2004).

[25] T. Donovan, D.J. Manning, P.W. Phillips, S. Higham, T. Crawford, "The effect of feedback on performance in a fracture detection task," in *Medical Imaging*, *Vol. 5749* (International Society for Optics and Photonics, 2005), pp. 79-85.

[26] H.L. Kundel, "How to minimize perceptual error and maximize expertise in medical imaging," in *Medical Imaging*, *Vol. 6515* (International Society for Optics and Photonics, 2007), pp. 651508-651508-651511.

[27] B.S. Kelly, L.A. Rainford, S.P. Darcy, E.C. Kavanagh, R.J. Toomey, "The Development of Expertise in Radiology: In Chest Radiograph Interpretation, “Expert” Search Pattern May Predate “Expert” Levels of Diagnostic Accuracy for Pneumothorax Identification," Radiology **280**, 252-260 (2016).

[28] E.M. Kok, H. Jarodzka, A.B.H. de Bruin, H.A.N. BinAmir, S.G.F. Robben, J.J.G. van Merriënboer, "Systematic viewing in radiology: seeing more, missing less?," Advances in Health Sciences Education **21**, 189-205 (2016).

[29] H.L. Kundel, C.F. Nodine, E.F. Conant, S.P. Weinstein, "Holistic Component of Image Perception in Mammogram Interpretation: Gaze-tracking Study," Radiology **242**, 396-402 (2007).

[30] S. Voisin, F. Pinto, G. Morin-Ducote, K.B. Hudson, G.D. Tourassi, "Predicting diagnostic error in radiology via eye-tracking and image analytics: Preliminary investigation in mammography," Medical Physics **40**, 101906 (2013).

[31] R. Engbert, K. Mergenthaler, "Microsaccades are triggered by low retinal image slip," Proceedings of the National Academy of Sciences **103**, 7192-7197 (2006).

[32] D.G. Stephen, J. Anastas, "Fractal fluctuations in gaze speed visual search," Attention, Perception, & Psychophysics **73**, 666-677 (2011).

[33] F.T. Alamudun, H.-J. Yoon, K. Hudson, G. Morin-Ducote, G. Tourassi, 2015 (unpublished).

[34] M. Heath, K. Bowyer, D. Kopans, P. Kegelmeyer, R. Moore, K. Chang, S. Munishkumaran, "Current Status of the Digital Database for Screening Mammography," in *Digital Mammography: Nijmegen, 1998,* edited by N. Karssemeijer, M. Thijssen, J. Hendriks, L. van Erning (Springer Netherlands, Dordrecht, 1998), pp. 457-460.

[35] A.C.o.R.B.-R. Committee, A.C.o. Radiology, *Breast imaging reporting and data system*. (American College of Radiology, 1998).

[36] B.B. Mandelbrot, D.E. Passoja, A.J. Paullay, "Fractal character of fracture surfaces of metals," 1984).

[37] T. Smith, W. Marks, G. Lange, W. Sheriff, E. Neale, "A fractal analysis of cell images," Journal of neuroscience methods **27**, 173-180 (1989).

[38] J.W. Fuseler, J.P. Robichaux, H.I. Atiyah, A.F. Ramsdell, "Morphometric and fractal dimension analysis identifies early neoplastic changes in mammary epithelium of MMTV-cNeu mice," Anticancer research **34**, 1171-1177 (2014).

[39] D.W. Sims, E.J. Southall, N.E. Humphries, G.C. Hays, C.J. Bradshaw, J.W. Pitchford, A. James, M.Z. Ahmed, A.S. Brierley, M.A. Hindell, "Scaling laws of marine predator search behaviour," Nature **451**, 1098-1102 (2008).

[40] A.M. Reynolds, A.D. Smith, R. Menzel, U. Greggers, D.R. Reynolds, J.R. Riley, "Displaced honey bees perform optimal scale-free search flights," Ecology **88**, 1955-1961 (2007).

[41] F. Bartumeus, M.E. da Luz, G. Viswanathan, J. Catalan, "Animal search strategies: a quantitative random-walk analysis," Ecology **86**, 3078-3087 (2005).

[42] F. Hausdorff, "Dimension und äußeres Maß," Mathematische Annalen **79**, 157-179 (1918).

[43] J. Eng, "ROC analysis: web-based calculator for ROC curves," *Vol. 2016* (Johns Hopkins University Baltimore, 2006).

[44] D.S. Al Mousa, P.C. Brennan, E.A. Ryan, W.B. Lee, J. Tan, C. Mello-Thoms, "How Mammographic Breast Density Affects Radiologists' Visual Search Patterns," Academic Radiology **21**, 1386-1393 (2014).

[45] E.A. Krupinski, A.A. Tillack, L. Richter, J.T. Henderson, A.K. Bhattacharyya, K.M. Scott, A.R. Graham, M.R. Descour, J.R. Davis, R.S. Weinstein, "Eye-movement study and human performance using telepathology virtual slides. Implications for medical education and differences with experience," Human Pathology **37**, 1543-1556 (2006).

[46] E.A. Krupinski, A.R. Graham, R.S. Weinstein, "Characterizing the development of visual search expertise in pathology residents viewing whole slide images," Human Pathology **44**, 357-364 (2013).

**APPENDIX A**

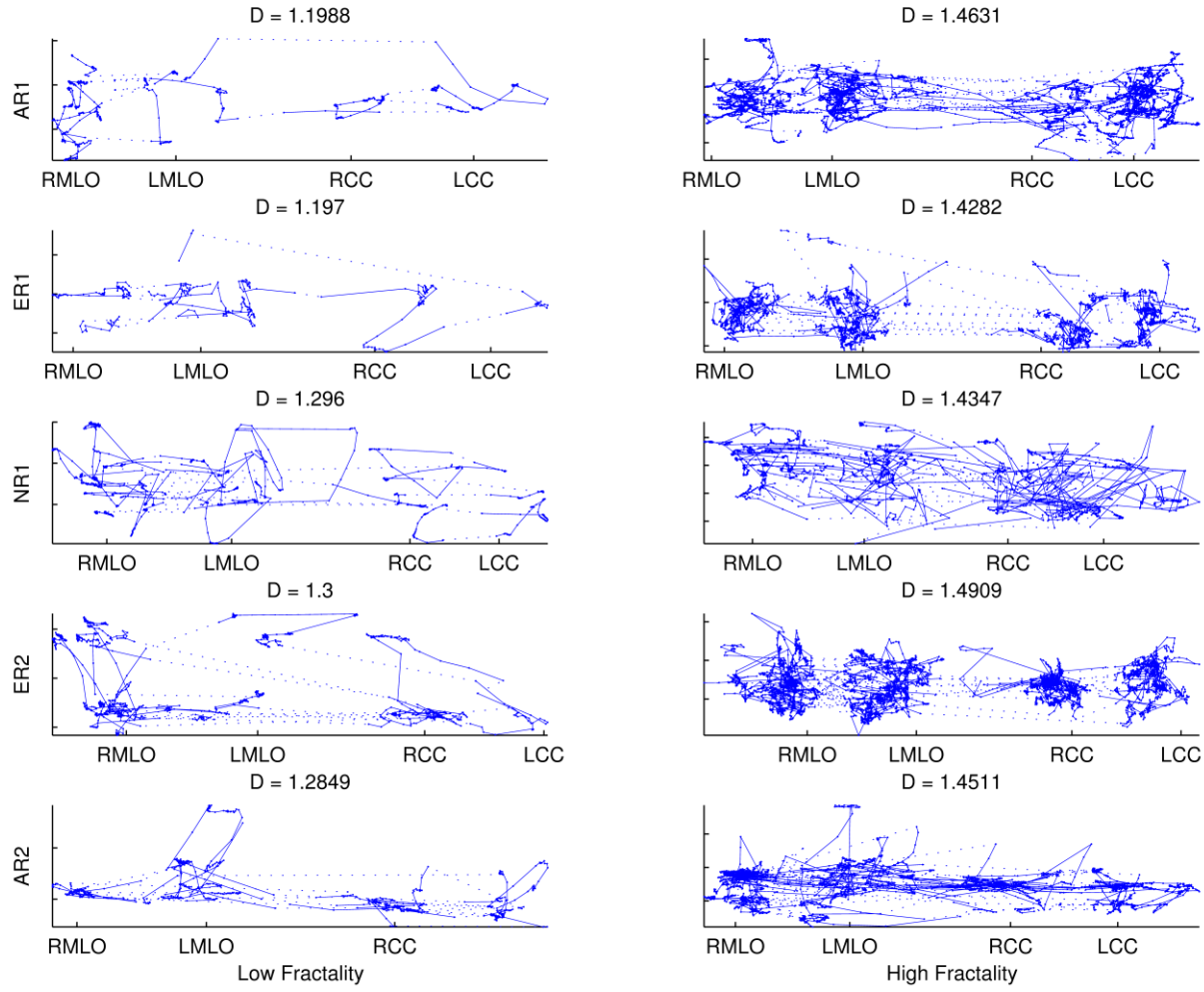



Figure 4. Illustration showing scanpaths with low (left) and high (right) fractal dimension for new Radiology residents (NR), advanced Radiology residents (AR), and expert radiologists (ER).

565

**APPENDIX B**

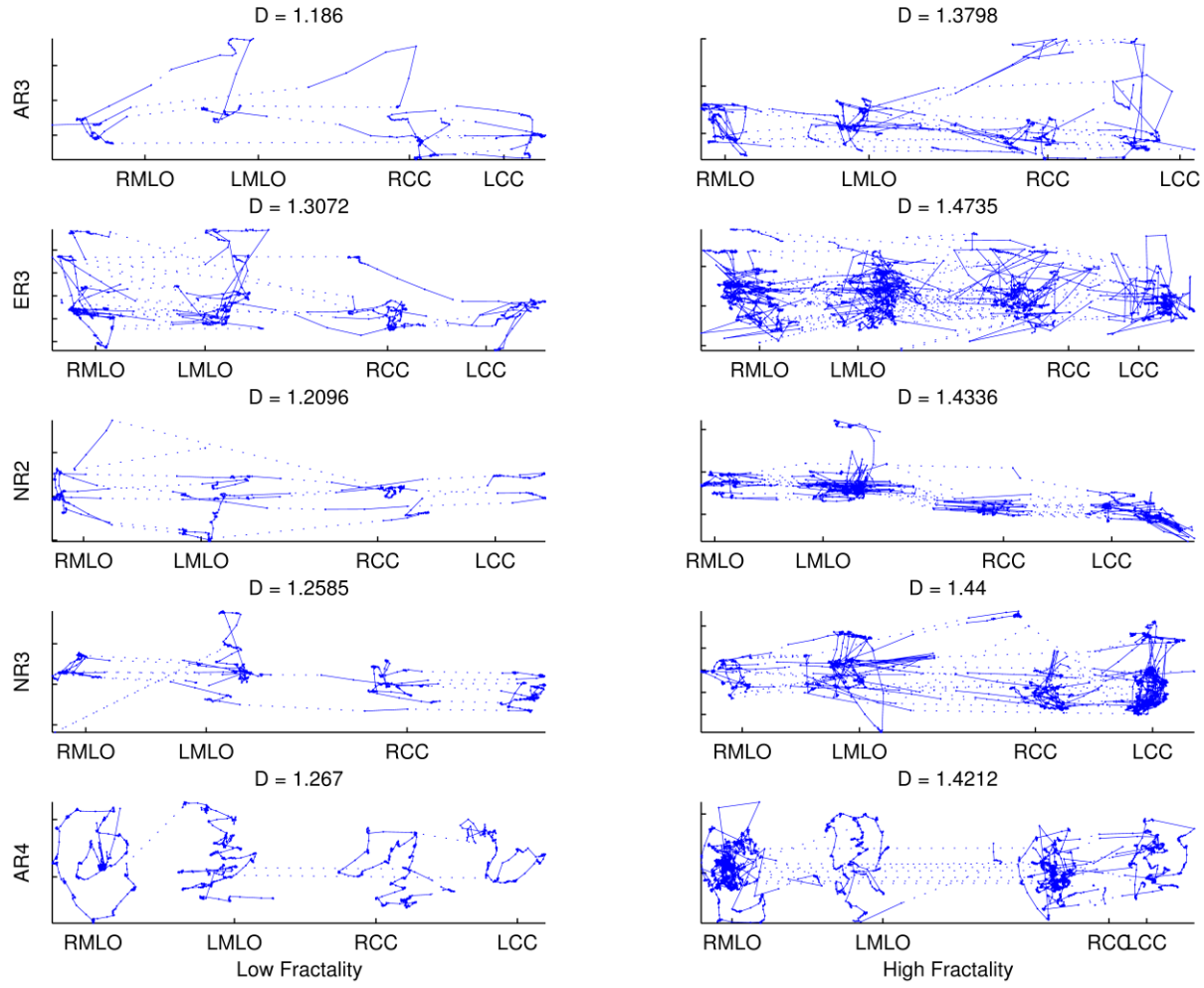



Figure 5. Illustration showing scanpaths with low (left) and high (right) fractal dimension for new Radiology residents (NR), advanced Radiology residents (AR), and expert radiologists (ER).

## APPENDIX C

The following table gives the time spent on each view, breast, and case averaged over all cases for each image reader. In addition, the table shows the average time each reader spent per case on the default configuration (i.e. the default 4-view configuration C6 from Table 4, RMLO|LMLO|RCC|LCC).

| | Average viewing time (s) | | | | | |
|---|---|---|---|---|---|---|
| **Reader** | **Cranial - Caudal** | **Mediolateral – Oblique** | **Right Breast** | **Left Breast** | **Default** | **Case** |
| NR1 | 4.87 ± 2.3 | 6.02 ± 2.4 | 4.02 ± 2.7 | 3.64 ± 2.0 | 3.58 ± 3.0 | 22.08 ± 9.3 |
| NR2 | 0.00 ± 0.0 | 0.00 ± 0.0 | 18.07 ± 0 | 9.08 ± 0.0 | 24.57 ± 14.2 | 24.84 ± 15.0 |
| NR3 | 7.51 ± 6.3 | 7.36 ± 4.0 | 11.20 ± 9.6 | 10.14 ± 9.0 | 16.63 ± 6.8 | 34.89 ± 14.9 |
| AR1 | 0.00 ± 0.0 | 0.00 ± 0.0 | 21.57 ± 16.3 | 20.38 ± 15.9 | 20.95 ± 11.4 | 38.19 ± 27.2 |
| AR2 | 3.96 ± 4.2 | 3.31 ± 2.4 | 3.80 ± 2.8 | 3.73 ± 1.9 | 12.22 ± 6.5 | 22.38 ± 10.3 |
| AR3 | 3.53 ± 1.7 | 3.50 ± 2.2 | 3.36 ± 1.9 | 3.63 ± 2.1 | 6.88 ± 3.8 | 17.67 ± 7.8 |
| AR4 | 24.66 ± 8.0 | 21.03 ± 8.8 | 21.06 ± 10.1 | 10.32 ± 3.5 | 2.73 ± 5.2 | 49.65 ± 14.7 |
| E1 | 3.99 ± 2.2 | 4.21 ± 2.7 | 4.40 ± 2.8 | 4.36 ± 2.9 | 5.41 ± 3.6 | 17.46 ± 11.4 |
| E2 | 5.16 ± 3.6 | 4.49 ± 2.8 | 3.73 ± 2.2 | 4.37 ± 3.0 | 14.49 ± 8.2 | 29.42 ± 13.4 |
| E3 | 7.47 ± 6.6 | 7.99 ± 6.4 | 7.10 ± 7.5 | 6.12 ± 6.5 | 13.08 ± 9.7 | 41.10 ± 26.5 |

570

## APPENDIX D

The following table gives the reference number of DDSM cases of malignant pathology used in our study.

| **Malignant** | | | | | |
|---|---|---|---|---|---|
| **volume** | **case number** | **breast density** | **abnormalities** | **mass shape** | **mass margin** |
| cancer 01 | case0001 | fibroglandular | 1 | IRR | SPIC |
| cancer 01 | case0003 | fibroglandular | 1 | IRR | SPIC |
| cancer 01 | case0004 | heterogeneous | 1 | IRR | SPIC |
| cancer 01 | case0006 | fibroglandular | 2 | IRR, ARCH | SPIC, N/A |
| cancer 01 | case0014 | fatty | 1 | IRR | ML |
| cancer 01 | case0016 | fatty | 1 | IRR | SPIC |
| cancer 01 | case0017 | dense | 1 | LOB | IDEF |
| cancer 01 | case3010 | fatty | 2 | IRR, IRR | IDEF, IDEF |
| cancer 01 | case3012 | heterogeneous | 1 | IRR | SPIC |
| cancer 01 | case3018 | fatty | 1 | LOB | IDEF |
| cancer 01 | case3022 | fatty | 1 | IRR | SPIC |
| cancer 01 | case3033 | heterogeneous | 1 | IRR | IDEF |
| cancer 01 | case3057 | heterogeneous | 1 | IRR | SPIC, MLOB |
| cancer 01 | case3073 | heterogeneous | 1 | IRR | IDEF |
| cancer 02 | case0018 | fibroglandular | 1 | LOB | MLOB |
| cancer 02 | case0027 | fatty | 1 | LOB | MLOB |
| cancer 02 | case0032 | heterogeneous | 1 | ARCH | IDEF |
| cancer 02 | case0034 | fibroglandular | 1 | IRR | SPIC |
| cancer 02 | case0035 | fibroglandular | 1 | OV | MLOB |
| cancer 02 | case0038 | fibroglandular | 2 | OV, FLB | CIRC, CLST |
| cancer 02 | case0040 | fatty | 1 | LOB | SPIC |
| cancer 02 | case0041 | fibroglandular | 1 | IRR | SPIC |
| cancer 02 | case0042 | fatty | 1 | IRR | SPIC |
| cancer 02 | case0043 | fibroglandular | 1 | IRR | MLOB |
| cancer 02 | case0059 | fibroglandular | 1 | IRR | SPIC |
| cancer 02 | case0070 | heterogeneous | 1 | ARCH | IDEF |
| cancer 02 | case0073 | fatty | 1 | ARCH | SPIC |
| cancer 02 | case0082 | fatty | 1 | LOB | CIRC |
| cancer 02 | case0089 | fibroglandular | 1 | IRR | SPIC |
| cancer 01 | case3023 | heterogeneous | 1 | ARCH | SPIC |
| cancer 05 | case0031 | fatty | 1 | IRR | MLOB |
| cancer 05 | case0085 | fatty | 1 | IRR | SPIC |
| cancer 05 | case0128 | fibroglandular | 1 | OV | MLOB |
| cancer 05 | case0140 | fatty | 1 | OV | MLOB |
| cancer 05 | case0142 | dense | 1 | IRR | SPIC |
| cancer 05 | case0143 | fatty | 1 | RND | SPIC |
| cancer 05 | case0146 | fibroglandular | 3 | OV, OV, OV | CIRC, CIRC, CIRC |
| cancer 05 | case0148 | heterogeneous | 1 | LOB, | IDEF |
| cancer 05 | case0149 | fibroglandular | 1 | OV | OB |
| cancer 05 | case0155 | fatty | 1 | LOB | SPIC |
| cancer 05 | case0156 | fatty | 1 | IRR | SPIC |
| cancer 05 | case0157 | fatty | 3 | OV, OV, OV | MLOB, MLOB, MLOB |
| cancer 05 | case0158 | fibroglandular | 1 | IRR | SPIC |
| cancer 05 | case0160 | fatty | 1 | LOB | MLOB |
| cancer 05 | case0161 | fibroglandular | 1 | LOB | MLOB |
| cancer 05 | case0164 | fatty | 1 | OV | CIRC |
| cancer 05 | case0165 | fibroglandular | 1 | RND | SPIC |
| cancer 05 | case0168 | fibroglandular | 1 | OV | OB |
| cancer 05 | case0170 | fibroglandular | 1 | OV | SPIC |
| cancer 05 | case0175 | heterogeneous | 1 | RND | SPIC |

IRR: Irregular,SPIC: spiculated, ARCH: architectural distortion, MLOB: microlobulated, LOB: Lobulated, IDEF: ill-defined, OV: oval, OVU: ovulated, FLB: fine linear branching, CIRC: circumscribed, CLST: clustered, RND: round, OB: obscured, AMPH: amorphous.

## APPENDIX E

The following table gives the reference number of the DDSM cases of benign pathology used in our study.

| **Benign** | | | | | |
|---|---|---|---|---|---|
| **volume** | **case number** | **breast density** | **total abnormalities** | **mass shape** | **mass margin** |
| benign 01 | case0217 | fibroglandular | 1 | RND | CIRC |
| benign 01 | case0240 | fibroglandular | 1 | OV | CIRC |
| benign 01 | case0243 | heterogeneous | 1 | OV | MLOB |
| benign 01 | case0245 | fibroglandular | 1 | OV | CIRC |
| benign 01 | case0248 | fibroglandular | 3 | RND, RND, LOB | CIRC, CIRC, MLOB |
| benign 01 | case0249 | fatty | 2 | LOB, LOB | CIRC, CIRC |
| benign 01 | case3093 | heterogeneous | 1 | ARCH | SPIC |
| benign 01 | case3098 | heterogeneous | 1 | IRR | IDEF |
| benign 01 | case3099 | fibroglandular | 1 | IRR | IDEF |
| benign 01 | case3100 | fibroglandular | 1 | RND | MLOB |
| benign 01 | case3113 | heterogeneous | 1 | RND | CIRC |
| benign 01 | case3118 | heterogeneous | 1 | RND | CIRC |
| benign 01 | case3128 | heterogeneous | 1 | IRR | SPIC |
| benign 01 | case3132 | heterogeneous | 1 | OVU | CIRC |
| benign 01 | case3140 | fibroglandular | 1 | RND | CIRC |
| benign 04 | case0251 | fibroglandular | 1 | IRR | IDEF |
| benign 04 | case0252 | fibroglandular | 1 | OV | CIRC |
| benign 04 | case0253 | fatty | 1 | OV | CIRC |
| benign 04 | case0273 | fibroglandular | 1 | RND | CIRC |
| benign 04 | case0274 | fibroglandular | 1 | OV | CIRC |
| benign 04 | case0282 | heterogeneous | 1 | OV | CIRC |
| benign 04 | case0283 | fibroglandular | 1 | ARCH | SPIC |
| benign 04 | case0303 | fatty | 1 | LOB | CIRC |
| benign 04 | case0304 | fibroglandular | 3 | LOB, AMPH, OV | CIRC, CLST, OB |
| benign 04 | case0306 | heterogeneous | 2 | LOB, LOB | OB, CIRC |

IRR: Irregular,SPIC: spiculated, ARCH: architectural distortion, MLOB: microlobulated, LOB: Lobulated, IDEF: ill-defined, OV: oval, OVU: ovulated, FLB: fine linear branching, CIRC: circumscribed, CLST: clustered, RND: round, OB: obscured, AMPH: amorphous.

## APPENDIX F

The following table gives the reference number of the DDSM cases of normal pathology used in our study.

| **Normal** | | |
|---|---|---|
| **volume** | **case number** | **breast density** |
| normal 09 | case3601 | fibroglandular |
| normal 09 | case3602 | fibroglandular |
| normal 09 | case3603 | fatty |
| normal 09 | case3604 | fibroglandular |
| normal 09 | case3606 | dense |
| normal 09 | case3607 | fibroglandular |
| normal 09 | case3608 | fibroglandular |
| normal 09 | case3609 | fibroglandular |
| normal 09 | case3611 | heterogeneous |
| normal 09 | case3612 | heterogeneous |
| normal 09 | case3613 | heterogeneous |
| normal 09 | case3615 | fibroglandular |
| normal 09 | case3618 | fatty |
| normal 09 | case3619 | heterogeneous |
| normal 09 | case3621 | fibroglandular |
| normal 10 | case3660 | fibroglandular |
| normal 10 | case3661 | fibroglandular |
| normal 10 | case3662 | dense |
| normal 10 | case3663 | fibroglandular |
| normal 10 | case3664 | fibroglandular |
| normal 10 | case3665 | heterogeneous |
| normal 10 | case3666 | fibroglandular |
| normal 10 | case3667 | fibroglandular |
| normal 10 | case3668 | fatty |
| normal 10 | case3670 | fibroglandular |

585