## *Mammographic Risk Assessment Based on Anatomical Linear Structures*

Hadley, Edward Michael; Denton, Erika R. E.; Zwiggelaar, Reyer

*Publication date:*
2006

*Citation for published version (APA):*
Hadley, E. M., Denton, E. R. E., & Zwiggelaar, R. (2006). *Mammographic Risk Assessment Based on Anatomical Linear Structures: Do Mammographic Abnormalities Play a Role?* . 11-15. Paper presented at 10th UK Conference on Medical Image Understanding and Analysis - MIUA, Manchester, United Kingdom of Great Britain and Northern Ireland.

# Mammographic Risk Assessment Based on Anatomical Linear Structures: Do Mammographic Abnormalities Play a Role?

Edward M. Hadley[a], Erika R. E. Denton[b] and Reyer Zwiggelaar[a*]

[a]Department of Computer Science, University of Wales, Aberystwyth, UK,
[b]Department of Radiology, Norfolk and Norwich University Hospital, UK

**Abstract.** Mammographic risk assessment is concerned with the probability of a woman developing breast cancer. Recently, it has been suggested that the density of linear structures is related to risk. For 321 images from the MIAS database, a measure of line strength was obtained for each pixel using the Line Operator method. The proportion of pixels with line strength above a threshold level was calculated for each image and the results categorised by Tabar pattern and Boyd SCC class, and the differences between results for the set of all images and the subset of 206 normal images were investigated. The results indicated a significant difference between Boyd classes 1–3 (low risk) and classes 4–6 (high risk), and between most Tabar patterns. No significant differences were found between the results for the set of all images and the set of normal images.

## 1 Introduction

Mammographic risk assessment is concerned with estimating the probability of women developing breast cancer. Risk assessment is a rapidly developing area of research and aims to improve the likelihood of the early detection of breast cancer. Breast density is an important indicator of mammographic risk [1] and the best predictor of mammographic sensitivity [2]. However, more recently, it has been suggested that the distribution of linear structures is also correlated with mammographic risk [3–5]. So far it is not entirely clear if it is just the density of linear structures (either by percentage area or volume) or if the distribution of the linear structures plays a role as well.

Tabar et al. have proposed a mammographic risk assessment model based on four structural components, where the relative proportions of each component is linked to the risk of developing breast cancer [3–5]. One of the four structural components is linear density. The main purpose of this work is to investigate if automatic methods can be used to correlate the density of linear structures to mammographic risk classification metrics.

Two classification models are used: Tabar patterns [5] and Boyd SCC classes [6]. Tabar's classification consists of five *patterns*, where patterns I–III represent a low risk of developing breast cancer, and patterns IV–V indicate a higher risk. Screening tests have shown that cancer prevalence in women with patterns IV-V is approximately twice that in women with patterns I-III [5]. The Boyd SCC model consists of a scale of six classes where class 1 indicates the lowest risk and class 6 indicates the highest risk. Since the Boyd model is based on density, correlation with the Boyd model can be seen as an indirect correlation with breast density. BIRADS classes [7] were also used in a more detailed anlysis of the initial results [8].

## 2 Method

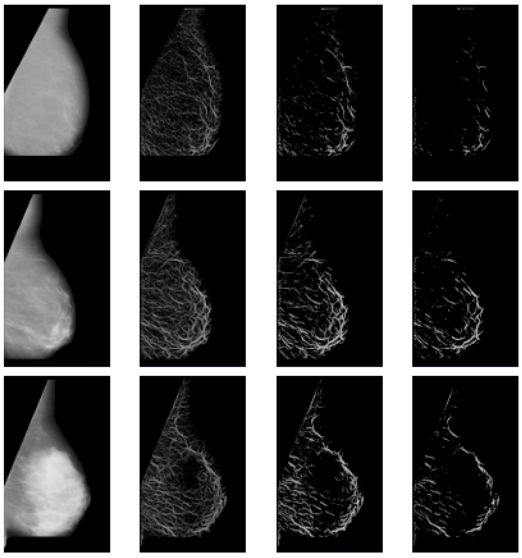
Three hundred and twenty-one mammographic images from the Mammographic Image Analysis Society (MIAS) database were classified according to Tabar patterns [5] and Boyd SCC classes [6] by an expert radiologist. Example images of low, moderate and high risk mammograms are shown in Fig. 1 (a).

The images were processed using Dixon and Taylor's line operator method [9, 10] (see Sect. 2.1), producing a measurement of line strength at each pixel. The multi-scale line operator has been shown to be more effective at detecting linear structures in mammographic images than other methods [10].

Figure 1 (b) shows examples of low, moderate and high risk mammograms following processing with the line operator.

The relative proportion of pixels with line strength values above a range of thresholds was calculated for each image. Figure 1 (c, d) shows examples of the resultant images after thresholding. Subsequently, the results were analysed for differences between images of each Tabar pattern and Boyd SCC class.

*Edward Hadley: emh05@aber.ac.uk, Reyer Zwiggelaar: rrz@aber.ac.uk.



(a) (b) (c) (d)

**Figure 1.** The top row shows a mammogram of Boyd SCC class 1/Tabar pattern II (low risk), the middle row shows a mammogram of Boyd SCC class 3/Tabar pattern III (moderate risk) and the bottom row shows a mammogram of Boyd SCC class 6/Tabar pattern IV (high risk). The images in column (a) show the original mammograms, column (b) shows the results after processing with the line operator, and columns (c) and (d) show the results after thresholding at 4/204 and 6/204 respectively. The lines in (b), (c) and (d) have been enhanced for viewing.

Following this analysis, the subset of normal (non-cancerous) images was selected and the results for this set compared with the results for the set of all images. The full set of 321 mammographic images contained 206 normal and 115 abnormal images. Graphs were produced showing a comparison of the two sets by Tabar pattern and Boyd class (see Figs. 4 and 5).

### 2.1 Line Operator

A study of various methods for detecting linear structures in mammograms [10] showed that Dixon and Taylor's line operator [9] is more accurate than other methods. As such, the line operator was used in our experiments. The method produces a measure of line strength and orientation for each pixel in an image.

The line orientation is determined by calculating the mean pixel brightness of a line of pixels running through the target pixel at a range of orientations. The orientation with the largest mean brightness is taken to be the line orientation. The line strength, $S$, is then given by

$$S = (L - N), \tag{1}$$

where $L$ is the mean brightness of the line of pixels, and $N$ is the mean brightness of a similarly orientated square of pixels.

Our experiment used a line length of five pixels and twelve orientations as suggested by earlier work [10].

A multi-scale approach was used in order to detect lines of a range of thicknesses and the resultant images were combined to produce line strength values for pixels at the original scale. Scaling of the images was achieved firstly by blurring the image using a 3x3 Gaussian kernel and subsequently by subsampling to provide a resultant image of half the width and height of the original. Our approach comprised processing with the line operator at three scales, since this appeared to produce the most reasonable output for the images under examination.

Finally, the pixel line strengths were thresholded to remove background texture. Using a line length of 5, the measures of line strength fall in the theoretical range of $0 - 204$, however the results showed that most (if not all) pixels had line strength values in the range $0 - 30$. A range of threshold values were chosen experimentally, and two values (4/204 and 6/204) were used for our analysis as they removed most background noise whilst maintaining most of the linear structure information (see Fig. 1 (c, d)).

## 3 Results

Initial results focussed on the set of all images. The relative above threshold linearities for the various Tabar patterns and Boyd SCC classes are shown in Figs. 2 and 3 respectively. These graphs provide an overview of the difference between the patterns and classes, and for more detailed analysis, Mann-Whitney tests were performed on each pair of Tabar patterns and Boyd SCC classes. These are shown in Tables 1 and 2 respectively. Parametric tests, such as analysis of variance (ANOVA) tests were not used because the test data did not fulfill the necessary assumptions, however the relatively small number of classes made pairwise Mann-Whitney tests possible.

The Mann-Whitney test results provide an indication as to whether there was a statistically significant difference between two classes of images in our data set. A significant difference would mean that it is possible to reliably distinguish between different classes of images.

As mentioned, two threshold values were used in our analysis (4/204 and 6/204). This is to demonstrate the effects of varying the threshold level and thus including more or less of the linear structure information.

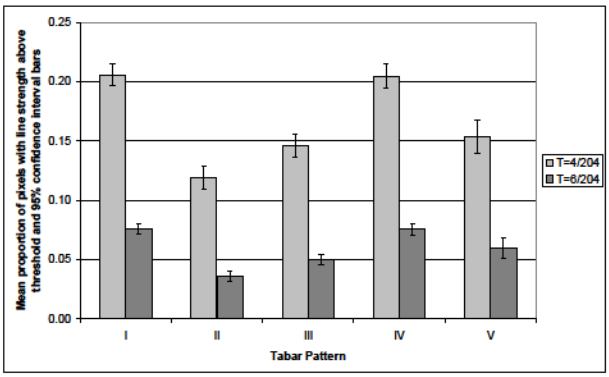



**Figure 2.** Graph showing the mean proportion of pixels with line strengths above each threshold $T$ and 95% confidence intervals for images of each **Tabar** pattern.

| Tabar | II | III | IV | V |
|---|---|---|---|---|
| I | 0.0000 | 0.0000 | 0.7938 | 0.0000 |
| II | | 0.0003 | 0.0000 | 0.0002 |
| III | | | 0.0000 | 0.3597 |
| IV | | | | 0.0000 |

$T = 4/204$

| Tabar | II | III | IV | V |
|---|---|---|---|---|
| I | 0.0000 | 0.0000 | 0.9069 | 0.0059 |
| II | | 0.0001 | 0.0000 | 0.0000 |
| III | | | 0.0000 | 0.0336 |
| IV | | | | 0.0050 |

$T = 6/204$

**Table 1.** The p-values obtained by Mann-Whitney tests on each combination of **Tabar** patterns for each threshold $T$. Results not significant at $\alpha = 0.05$ are shaded.

The results of the analysis by Tabar pattern (see Fig. 2, Table 1) demonstrate the ability to reliably distinguish between patterns II–III (low risk) and patterns IV–V (high risk) at a threshold of 6/204, with low risk pattern I being indistinguishable from high risk pattern IV. The results at a threshold of 4/204 are less promising, since the low risk pattern III becomes indistinguishable from high risk pattern V.

The results of the analysis by Boyd SCC class (see Fig. 3, Table 2) differ somewhat, and it is clear that the proposed method is able to distinguish between classes 1–3 and classes 4–6 and both thresholds, with a general trend through classes 1–4/5 indicating that a greater linear density is indicative of a greater risk. The three lower risk classes are each distinguishable from all other classes, whilst the three higher risk classes are distinguishable from the three lower risk classes, but are indistinguishable from one another at $T = 6/204$. At $T = 4/204$ classes 4 and 6 become distinguishable from one another.

Following initial analysis, we investigated whether removing the abnormal images from the data set had an effect on results. The results of this analysis by Tabar pattern and Boyd SCC class are show in Figs. 4 and 5, respectively. The results show that there is no significant difference between the above threshold linearity in normal and abnormal images of each pattern or class.

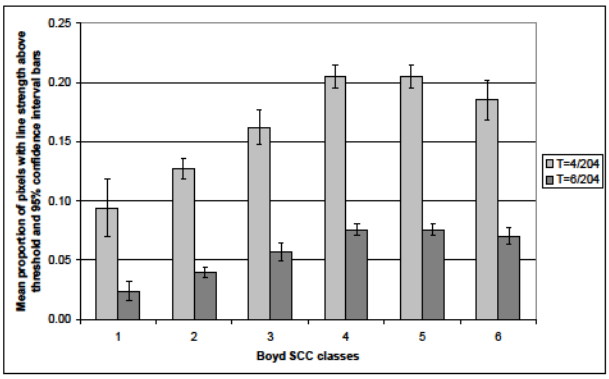



**Figure 3.** Graph showing the mean proportion of pixels with line strengths above each threshold $T$ and 95% confidence intervals for images of each **Boyd SCC** class.

| Boyd | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 1 | 0.0332 | 0.0023 | 0.0001 | 0.0001 | 0.0008 |
| 2 | | 0.0002 | 0.0000 | 0.0000 | 0.0000 |
| 3 | | | 0.0000 | 0.0000 | 0.0362 |
| 4 | | | | 0.6693 | 0.0287 |
| 5 | | | | | 0.0833 |

$T = 4/204$

| Boyd | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 1 | 0.0155 | 0.0009 | 0.0001 | 0.0000 | 0.0004 |
| 2 | | 0.0001 | 0.0000 | 0.0000 | 0.0000 |
| 3 | | | 0.0000 | 0.0000 | 0.0031 |
| 4 | | | | 0.7621 | 0.4828 |
| 5 | | | | | 0.6612 |

$T = 6/204$

**Table 2.** The p-values obtained by Mann-Whitney tests on each combination of **Boyd SCC** classes at each threshold $T$. Results not significant at $\alpha = 0.05$ are shaded.

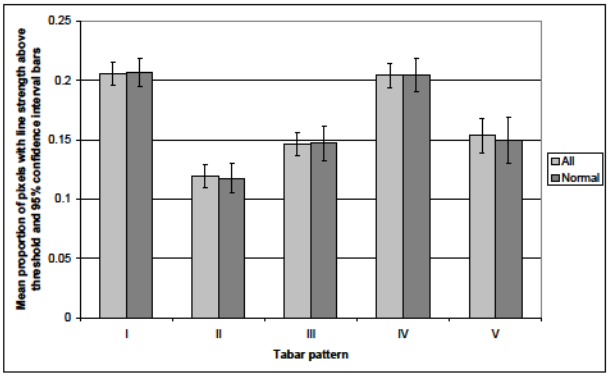



**Figure 4.** Graph showing the mean above threshold linearity and 95% confidence intervals for all images and the subset of normal images at each **Tabar** pattern.

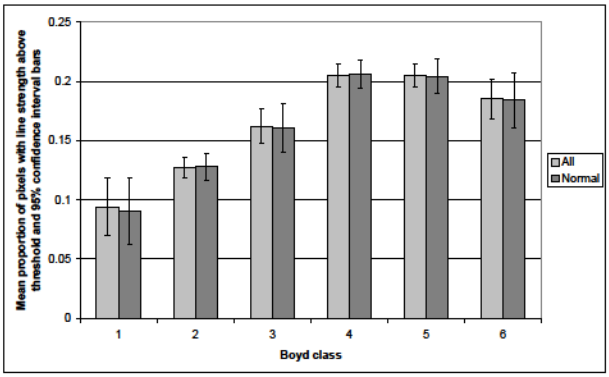



**Figure 5.** Graph showing the mean above threshold linearity and 95% confidence intervals for all images and the subset of normal images at each **Boyd** class.

## 4 Discussion and Conclusions

Whilst the proposed approach is simplistic, the results are promising and the analysis by Boyd class demonstrates a clear ability to automatically distinguish between lower risk (classes 1–3) and higher risk mammograms (classes 4–6).

The analysis by Tabar pattern is interesting. The results for pattern V were unexpected. Tabar suggests that the linear density in pattern V mammograms should be very low [5], whereas our results indicate a relatively high linear density for this pattern. However, a possible explanation might be that the line operator enhances linear structures even in dense tissue (see Fig. 1) and as such might result in a high proportion of linear structures, whereas under the Tabar classification this area might be assigned to one of the alternative mammographic building blocks, such as fibrous tissue. It is intended in future work to segment and mask the fibrous tissue prior to application of the line detector.

In addition, the results for Tabar pattern I mammograms, which demonstrate a high linear density, do not correlate with the results of low risk Boyd or BIRADS classes, which show low linear densities. Other studies have found that Tabar's patterns do not correlate well with other risk assessment models [11]. The principal anomaly is the low risk pattern I. The mammograms in our test set belonging to this pattern do not easily correlate with a particular Boyd class, instead being spread amongst Boyd classes 2-5, with the majority seeming to belong to the high risk classes 4-5 [11].

We see in the analysis by Boyd class that the highest class (6) shows a slightly lower linear density than the classes immediately below it. At a threshold of 6/204 this decline is only slight and there is no significant difference between the linear densities of classes 4–6 (see Table 2). The density of class 6 is also significantly higher than all three of the low risk classes. The decline is accentuated when a threshold of 4/204 is taken, where we see a significant difference between classes 4 and 6. This indicates a dependence on the threshold value and part of our future work will include an investigation in to a more principled approach to determine the threshold.

The results of a comparison between the whole data set and the subset of normal images shows no significant differences between the two sets. This is to be expected as the development of an abnormality is unlikely to increase or decrease the quantity of linear structures (with the exception of spiculations). It is possible, however, that it may affect the distribution of the linear structures.

In summary, the proposed approach is promising but simplistic in that it considers only the density of linear structures and does not take in to account information relating to their distribution. Since several risk assessment models are based on the parenchymal patterns in the breast [1, 3, 5], it is intended for further work to investigate whether the distribution of linear structures in addition to their density is related to mammographic risk, and whether this information can be used to improve risk assessment classification.

## References


1. J. N. Wolfe. "Risk for breast cancer development determined by mammographic parenchymal pattern." *Cancer* **37(5)**, pp. 2486–2492, 1976.
2. T. M. Kolb, J. Lichy & J. H. Newhouse. "Comparison of the performance of screening mammography, physical examination, and breast us and evaluation of factors that influence them: An analysis of 27,825 patient evaluations." *Radiology* **225(1)**, pp. 165–175, 2002.
3. L. Tabar & P. B. Dean. "Mammographic parenchymal patterns. risk indicator for breast cancer?" *Journal of the American Medical Association* **247(2)**, pp. 185–189, 1982.
4. I. T. Gram, E. Funkhouser & L. Tabar. "The Tabar classification of mammographic parenchymal patterns." *European Journal of Radiology* **24(2)**, pp. 131–136, 1997.
5. L. Tabar, T. Tot & P. B. Dean. *Breast Cancer - The Art and Science of Early Detection with Mammography.* Georg Thieme Verlag, Stuttgart, 2005.
6. N. F. Boyd, J. W. Byng, R. A. Jong et al. "Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian National Breast Screening Study." *Journal of the National Cancer Institute* **87**, pp. 670–675, 1995.
7. American College of Radiology. *Illustrated Breast Imaging Reporting and Data System.* American College of Radiology, third edition, 1998.
8. E. M. Hadley, E. R. E. Denton & R. Zwiggelaar. "Mammographic risk assessment based on anatomical linear structures." *Lecture Notes in Computer Science* **4046**, pp. 626–633, 2006.
9. R. N. Dixon & C. J. Taylor. "Automated asbestos fibre counting." *Institute of Physics Conference Series* **44**, pp. 178–185, 1979.
10. R. Zwiggelaar, S. M. Astley, C. R. M. Boggis et al. "Linear structures in mammographic images: Detection and classification." *IEEE Transactions on Medical Imaging* **23(9)**, pp. 1077–1086, 2004.
11. I. Muhimmah, A. Oliver, E. R. E. Denton et al. "Comparison between Wolfe, Boyd, BI-RADS and Tabar based mammographic risk assessment." *Lecture Notes in Computer Science* **4046**, 2006.