

# Comparison Between Wolfe, Boyd, BI-RADS and Tabár Based Mammographic Risk Assessment

Izzati Muhimmah<sup>1</sup>, Arnau Oliver<sup>2</sup>, Erika R.E. Denton<sup>3</sup>, Josep Pont<sup>4</sup>,  
Elsa Pérez<sup>4</sup>, and Reyer Zwiggelaar<sup>1</sup>

<sup>1</sup> Department of Computer Science, University of Wales Aberystwyth, UK  
`iim04@aber.ac.uk`, `rrz@aber.ac.uk`

<sup>2</sup> Dept. of Electronics, Informatics & Automatics, University of Girona, Spain

<sup>3</sup> Department of Breast Imaging, Norwich and Norfolk University Hospital, UK

<sup>4</sup> University Hospital Dr. Josep Trueta of Girona, Spain

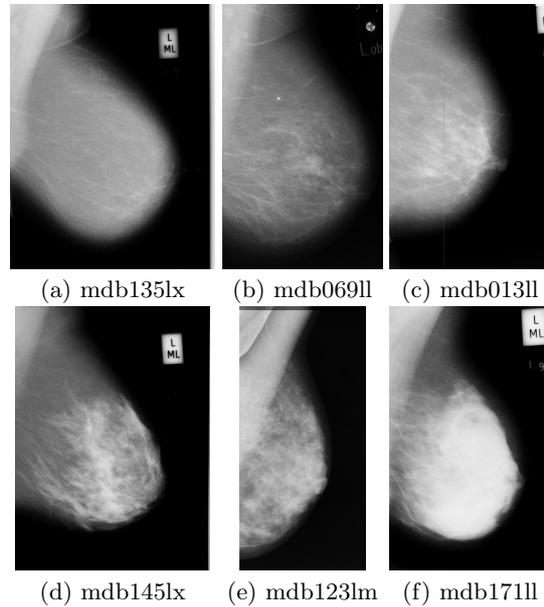
**Abstract.** Mammographic risk assessment provides an indication of the likelihood of women developing breast cancer. A number of mammographic image based classification methods have been developed, such as Wolfe, Boyd, BI-RADS and Tabár based assessment. We provide a comparative study of these four approaches. Results on the full MIAS database are presented, which indicate strong correlation (Spearman's  $> 0.9$ ) between Wolfe, Boyd and BI-RADS based classification, whilst the correlation with Tabár based classification is less straight forward (Spearman's  $< 0.5$ , but low correlations mainly caused by one of the classes).

## 1 Introduction

Mammographic risk assessment metrics commonly used are those based on Wolfe [1], Boyd [2], Tabár [3], or BI-RADS [4] (see Figure 1 for examples). These four metrics can be grouped into two approaches of assessment. Boyd's measures the percentage area of dense breast tissue. By way of contrast, Wolfe, BI-RADS, and Tabár all include patterns and texture information in estimating the classification. The main aim of this study is to investigate how these four metrics are correlated. Brisson *et al.* [5] studied correlation between Wolfe and Boyd metrics. Gram *et al.* [6] reported correlation between Tabár and Wolfe based classification on Tromsø screening mammograms. Gram *et al.* [7] reported a study about correlation between Wolfe, Boyd and Tabár metrics. To our knowledge, this is the first study to investigate the correlation between Wolfe, Boyd, Tabár and BI-RADS classification on a well known publicly available database [8].

### 1.1 Mammographic Risk Assessment Metrics

Mammographic risk assessment is often related to breast density estimation, and this is claimed to be a robust risk indicator. Moreover, Byrne *et al.* claimed that mammographic density is the strongest risk factor for breast cancer [9]. It should be noted that density estimation can also be used to evaluate how likely abnormalities are hidden from the observer [10].



**Fig. 1.** Example mammograms, where: (a) SCC: 0%, W: N1, T: Pattern II, B: I (b) SCC: 0 – 10%, W: N1, T: Pattern III, B: I (c) SCC: 11 – 25%, W: P1, T: Pattern III, B: II (d) SCC: 26 – 50%, W: P2, T: Pattern I, B: III (e) SCC: 51 – 75%, W: P2, T: Pattern IV, B: III and (f) SCC: > 75%, W: DY, T: Pattern V, B: IV.

Wolfe [1] proposed four categories of mammographic risk: N1 is defined as a mammogram that is composed mainly of fat and a few fibrous tissue strands; P1 shows a prominent duct pattern and a beaded appearance can be found either in the subareolar area or the upper axillary quadrant; P2 indicates severe involvement of a prominent duct pattern which may occupy from one-half up to all of the volume of the parenchyma and often the connective tissue hyperplasia produces coalescence of ducts in some areas; DY features a general increase in density of the parenchyma (which might be homogeneous) and there may or maynot be a minor component of prominent ducts. These four groups had an incidence of developing breast cancer of 0.1, 0.4, 1.7 and 2.2, respectively [1].

Boyd *et al.* [2] introduced a quantitative classification of mammographic densities. It is based on the proportion of dense breast tissue relative to the breast areas. The classification is known as Six Class Categories (SCC) where the density proportions are: Class1: 0%, Class2: (0 – 10%), Class3: [10 – 25%), Class4: [25 – 50%), Class5: [50 – 75%), and Class6: [75 – 100%). The increase in the level of breast tissue density has been associated with increase in the risk of developing breast cancer, specifically the relative risk for SCC 3 to 6 are 1.9, 2.2, 4.6, and 7.1, respectively [2].

Tabár *et al.* [3] describes breast composition of four building blocks: nodular density, linear density, homogeneous fibrous tissue, and radiolucent adipose tissues which also define mammographic risk classification. Pattern I: mammograms

are composed of 25%, 16%, 35%, and 24% of the four building blocks, respectively; Pattern II has approximate compositions as: 2%, 14%, 2%, and 82%; Pattern III is quite similar in composition to Pattern II, except that the retroareolar prominent ducts are often associated with periductal fibrosis; Pattern IV is dominated by prominent nodular and linear densities, with compositions of 49%, 19%, 15%, and 17%; Pattern V is dominated by extensive fibrosis and is composed as 2%, 2%, 89%, and 7% of the building blocks. Tabár *et al.* defined Patterns I-III corresponding to lower breast cancer risk, whilst Patterns IV-V relate to higher risk [3].

There are four BI-RADS [4] categories, which are: BI-RADS I: the breast is almost entirely fatty; BI-RADS II: there is some fibroglandular tissue; BI-RADS III: the breast is heterogeneously dense; BI-RADS IV: the breast is extremely dense. Lam *et al.* reported associations between BI-RAD II-IV and breast carcinoma (adjusted for weight) in postmenopausal women which were 1.6, 2.3, and 4.5, respectively [11].

## 2 Material and Methods

To investigate the correlation between the four mammographic risk assessment metrics, 321 images (case mdb29511 has not been included for historical reasons) from the MIAS database [8] were classified by three experienced breast screening radiologists (ED, JP, ES). All the mammograms were digitised (8-bits) with a scanning microdensitometer (Joyce-Loebl, SCANDIG3) to 50 micron  $\times$  50 micron resolution. The grey-scale response of the instrument is linear in the optical density range 0-3.2OD [8]. It should be noted that the mammograms were displayed on a standard PC monitor, which cannot be used for diagnostic purposes but is sufficient for mammographic risk assessment.

All results are shown in the form of confusion matrices. We have also computed the Spearman's correlation ( $r_S$ ) between the metrics (using SPSS version 13 for Windows) and linear-weighted kappa values ( $\kappa$ ) [12] (it should be noted that  $\kappa$  only tends to make sense when an equal number of classes is compared, but  $\kappa$  is provided for all cases for completeness).

### 2.1 Correlation Between Metrics

This part of the evaluation is based on assessment by one (ED) of the expert radiologists. All 321 images were classified according to Wolfe (N1, P1, P2, and DY), Boyd (Class 1-6), Tabár (Pattern I-V), and BI-RADS (I-IV). The images were displayed according to MIAS's numbering. It should be noted that Tabár and BI-RADS methods are not routinely used by the radiologist and all classifications for each mammogram were obtained at the same time (both these aspects might introduce bias).

### 2.2 Intra and Inter Observer Variation

To address the reproducibility, we compared the radiologist (ED) ratings to the same radiologist's previous assessments of Wolfe and Boyd's SCC. It should

be noted that for this intra-observer results, the data was assessed twice with the initial assessments two years before those described in section 2.1. In addition, the number of cases for *Rating 1*  $n = 319$  and for *Rating 2*  $n = 320$ , which were due to a technical problem in displaying the cases mdb321lm and/or mdb322rm.

We also compared the BI-RADS ratings by one radiologist (ED) with assessment by two other experts (JP,EP). It should be noted that there was a slight difference in protocol for JP and EP in that images were presented in left-right pairs, instead of individual images as was the case for ED.

### 3 Results and Discussions

#### 3.1 Correlation Between Metrics

The confusion matrices for all assessment by radiologist (ED) are shown in Tables 1- 6.

**Table 1.** Expert radiologist (ED) classification according to Boyd and Wolfe

		Wolfe			
		N1	P1	P2	DY
Boyd	SCC1	6	0	0	0
	SCC2	55	5	0	0
	SCC3	1	44	1	0
	SCC4	0	41	34	0
	SCC5	0	2	72	16
	SCC6	0	0	0	44

Table 1 shows that Boyd's Class 1 and 6 are all grouped as Wolfe's N1 and DY, respectively. The distribution of Class 2-5 are mainly mapped into lower risk according to Wolfe. The correlation for these two measures was  $r_S = 0.928$  ( $\kappa = 0.2033$ ). This is in line with a study reported by Brisson *et al.* [5] which showed a correlation of  $r_S = 0.81$  ( $P = 0.0001$ ). Moreover, they concluded that Wolfe's classification was redundant when percentage density was available in breast cancer risk assessment, which is supported by the results presented in Table 1.

**Table 2.** Expert radiologist (ED) classification according to BI-RADS and Wolfe

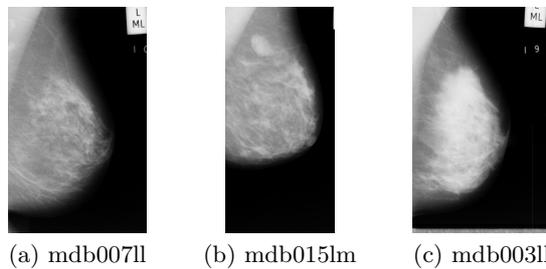
		Wolfe			
		N1	P1	P2	DY
BI-RADS	I	58	1	0	0
	II	4	80	2	0
	III	0	11	104	27
	IV	0	0	1	33

Table 2 shows a high agreement between Wolfe and BI-RADS measures, with a correlation of  $r_S = 0.929$  ( $\kappa = 0.8645$ ).

Table 3 shows that Tabár’s Pattern V is all grouped as Wolfe’s DY. The correlation for these two measures was  $r_S = 0.454$  ( $\kappa = 0.204$ ). By excluding Pattern I, Tabár and Wolfe show high correlation  $r_S = 0.93$  ( $\kappa = 0.8378$ ). Gram *et al.* reported result on this agreement was  $\kappa = 0.23$  [6]. They also showed that Tabár’s Pattern I corresponds to Wolfe’s DY in 45.6% of the mammograms and Pattern II to V has a unique mapping into Wolfe N1 to DY, respectively [6]. The recently published study by Gram *et al.* [7] reported moderate agreement between Wolfe and Tabár metric ( $\kappa = 0.51$ ) and here the mappings between Tabár and Wolfe based classifications were similar to our result. Some examples of images which have Tabár’s Pattern I and various Wolfe’s classes can be seen in Figure 2, which clearly shows the variation for Tabár’s Pattern I.

**Table 3.** Expert radiologist (ED) classification according to Tabár and Wolfe

		Wolfe			
		N1	P1	P2	DY
Tabár	I	0	61	56	2
	II	52	1	0	0
	III	10	30	0	0
	IV	0	0	51	30
	V	0	0	0	28



**Fig. 2.** Example mammograms which were rated as Tabár’s Pattern I and various Wolfe’s classes: (a) P1, (b) P2, (c) DY

Table 4 shows the agreement between BI-RADS and Tabár measures, with a correlation of  $r_S = 0.408$  ( $\kappa = 0.1347$ ). However, as shown above, when ignoring the Tabár’s Pattern I results the correlation increases to  $r_S = 0.96$  ( $\kappa = 0.9145$ ).

Table 5 shows a high agreement between BI-RADS and Boyd measures, with a correlation of  $r_S = 0.908$  ( $\kappa = 0.1792$ ).

Table 6 shows agreement between Boyd and Tabár measures, with a correlation of  $r_S = 0.459$  ( $\kappa = 0.2127$ ). However, as shown above, when excluding the Tabár’s Pattern I results the correlation increases to  $r_S = 0.93$  ( $\kappa = 0.5679$ ).

**Table 4.** Expert radiologist (ED) classification according to Tabár and BI-RADS

		BI-RADS			
		I	II	III	IV
Tabár	I	0	54	65	0
	II	50	3	0	0
	III	9	29	2	0
	IV	0	0	75	6
	V	0	0	0	28

**Table 5.** Expert radiologist (ED) classification according to Boyd and BI-RADS

		BI-RADS			
		I	II	III	IV
Boyd	SCC1	6	0	0	0
	SCC2	53	7	0	0
	SCC3	0	46	0	0
	SCC4	0	33	42	0
	SCC5	0	0	84	6
	SCC6	0	0	16	28

**Table 6.** Expert radiologist (ED) classification according to Boyd and Tabár

		Tabár				
		I	II	III	IV	V
Boyd	SCC1	0	6	0	0	0
	SCC2	1	45	14	0	0
	SCC3	21	2	23	0	0
	SCC4	68	0	3	4	0
	SCC5	29	0	0	60	1
	SCC6	0	0	0	17	27

**Table 7.** Spearman’s correlation between four different measures. Within brackets are the Spearman’s correlation when Tabár Pattern I is excluded.

	Boyd	Tabár	BI-RADS
Wolfe	0.928	0.454 (0.93)	0.929
Boyd		0.459 (0.93)	0.908
Tabár			0.408 (0.96)

Correlations are significant at the level of 0.01 (2-tailed).

A summary of the correlation between the four measures (from Tables 1- 6) can be found in Table 7. This shows that Wolfe - Boyd and Wolfe - BI-RADS have similar high correlation values, followed by the Boyd - BI-RADS correlation. It should be noted that such correlation does not necessarily imply that the metrics are based on the same information and this needs further investigation.

In contrast, Tabár’s does not correlate with the other measures. It is pointed out by Gram *et al.* that Tabár’s classification captures something more than just density measurements and its relation to breast cancer risk needs further investigation [7].

### 3.2 Intra Observer Variation

We present the intra-reproducibility of our radiologist (ED) on Wolfe’s and Boyd’s SCC metrics in Table 8 and Table 9, respectively. Intra-radiologist agreement on Wolfe’s classification were  $r_S = 0.81$  ( $\kappa = 0.5999$ ) and  $r_S = 0.85$  ( $\kappa = 0.6606$ ) for the two previous assessments compared to the most recent (*Rating 3*) assessment. For SCC, the intra-radiologist agreement were  $r_S = 0.89$  ( $\kappa = 0.6989$ ) and  $r_S = 0.90$  ( $\kappa = 0.7181$ ). These indicate a moderate to good agreement. It should be noted that for both metrics the most recent assessment shows a clear shift to higher risk classes when compared to previous assessment.

Part of our future research will concentrate on extending these intra-observer aspects.

**Table 8.** Intra-observer (ED) reproducibility for Wolfe based assessment

		Rating 3			
		N1	P1	P2	DY
Rating 1	N1	62	59	8	0
	P1	0	8	9	0
	P2	0	25	88	40
	DY	0	0	0	20

(a)  $\kappa = 0.5999$

		Rating 3			
		N1	P1	P2	DY
Rating 2	N1	62	60	5	0
	P1	0	17	18	0
	P2	0	15	81	25
	DY	0	0	2	35

(b)  $\kappa = 0.6606$

**Table 9.** Intra-observer (ED) reproducibility for Boyd’s SCC based assessment

		Rating 3					
		SCC1	SCC2	SCC3	SCC4	SCC5	SCC6
Rating 1	SCC1	3	8	0	0	0	0
	SCC2	3	48	13	3	0	0
	SCC3	0	4	28	25	10	0
	SCC4	0	0	5	44	35	0
	SCC5	0	0	0	1	45	27
	SCC6	0	0	0	0	0	17

(a)  $\kappa = 0.6989$

		Rating 3					
		SCC1	SCC2	SCC3	SCC4	SCC5	SCC6
Rating 2	SCC1	1	5	0	0	0	0
	SCC2	5	52	10	2	0	0
	SCC3	0	3	33	31	6	0
	SCC4	0	0	3	39	40	1
	SCC5	0	0	0	2	42	19
	SCC6	0	0	0	0	2	24

(b)  $\kappa = 0.7181$

### 3.3 Inter Observer Variation

To evaluate the inter-observer variations, we compared BI-RAD bases assessment by three radiologists. The results are presented in Table 10. The agreement between ED and two other radiologists were  $r_S = 0.85$  ( $\kappa = 0.5699$ ) and  $r_S = 0.82$  ( $\kappa = 0.6381$ ), respectively, whilst agreement between JP and

ES was  $r_S = 0.82$  ( $\kappa = 0.7139$ ). It should be noted the results of radiologist 1 (ED) tends toward higher BI-RADS classes when compared to the other radiologist.

Future research will concentrate on extending these inter-observer variation evaluation, ensuring we cover the full range of metrics and a similar protocol.

**Table 10.** Inter-observer reproducibility for BI-RADS based assessment

		Radiologist 1			
		I	II	III	IV
Rad. 2	I	57	66	6	0
	II	2	20	57	0
	III	0	0	62	7
	IV	0	0	17	27

(a)  $\kappa = 0.5699$

		Radiologist 1			
		I	II	III	IV
Rad. 3	I	57	29	0	0
	II	2	48	62	0
	III	0	9	64	7
	IV	0	0	16	27

(b)  $\kappa = 0.6381$

		Radiologist 2			
		I	II	III	IV
Rad. 3	I	83	3	0	0
	II	38	57	17	0
	III	8	19	46	7
	IV	0	0	6	37

(c)  $\kappa = 0.7139$

## 4 Conclusion

We have investigated the correlations between four different mammographic risk assessments on the MIAS database. The results show strong correlations among Wolfe/BI-RADS/Boyd metrics. However, Tabár based assessment is less correlated to the other three metrics. In addition, intra- and inter-observer variation results have been presented and discussed.

## Acknowledgments

I. Muhimmah would like to gratefully acknowledge funding from the Islamic Development Bank and the University of Wales Aberystwyth that enabled the presented research. A. Oliver would like to acknowledge MEC grant number TIN2005-08792-C03-01.

## References

1. Wolfe, J.N.: Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer* **37** (1976) 2486–2492
2. Boyd, N., Byng, J., Jong, R., Fishell, E., Little, L., Miller, A., Lockwood, G., Tritchler, D., Yaffe, M.J.: Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian national breast screening study. *Journal of the National Cancer Institute* **87** (1995) 670–675
3. Tabár, L., Tot, T., Dean, P.B.: *Breast Cancer: The Art and Science of Early Detection with Mammography*. Georg Thieme Verlag (2005)
4. ACR: *Breast Imaging Reporting and Data System (BI-RADS) 2<sup>nd</sup> edition*. Reston: American College of Radiology (1995)
5. Brisson, J., Diorio, C., Mâsse, B.: Wolfe's parenchymal pattern and percentage of the breast with mammographic densities: redundant or complementary classifications? *Cancer Epidemiol Biomarkers Prev* **12** (2003) 728–732

6. Gram, I.T., Funkhouser, E., Tabár, L.: The Tabár classification of mammographic parenchymal patterns. *European Journal of Radiology* **24** (1997) 131–136
7. Gram, I.T., Bremmes, Y., Ursin, G., Maskarinec, G., Bjurstam, N., Lund, E.: Percentage density, Wolfe's and Tabár's mammographic patterns: agreement and association with risk factors for breast cancer. *Breast Cancer Research* **7** (2005) R854–R861
8. Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S., Taylor, P., Betal, D., Savage, J.: The mammographic image analysis society digital mammogram database. *Excerpta Medica. International Congress Series* **1069** (1994) 375–378
9. Byrne, C., Schairer, C., Wolfe, J., Parekh, N., Salane, M., Brinton, L.A., Hoover, R., Haile, R.: Mammographic features and breast cancer risk: Effects with time, age, and menopause status. *Journal of National Cancer Institute* **87** (1995) 1622–1629
10. Egan, R., Mosteller, R.: Breast cancer mammography patterns. *Cancer* **40** (1977) 2087–2090
11. Lam, P.B., Vacek, P.M., Geller, B.M., Muss, H.B.: The association of increased weight, body mass index, and tissue density with the risk of breast carcinoma in Vermont. *Cancer* **89** (2000) 369–375
12. Lowry, R.: KAPPA Calculator [<http://faculty.vassar.edu/lowry/kappa.html>]. (Accessed 7/3/06)