# *Image Based Tissue Segmentation: Towards the Automation of Mammographic Risk Assessment*

Zwiggelaar, Reyer; Strange, Harry George; He, Wenda; Chen, Zhili; Kshirasagar, Ashwini; Denton, Erika R.E.

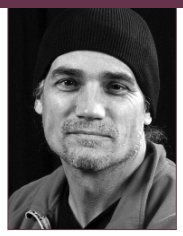
*Published in:*
Oncology News

*Publication date:*
2014

*Citation for published version (APA):*
Zwiggelaar, R., Strange, H. G., He, W., Chen, Z., Kshirasagar, A., & Denton, E. R. E. (2014). Image Based Tissue Segmentation: Towards the Automation of Mammographic Risk Assessment. *Oncology News*, *9*, 12-15. http://media.wix.com/ugd/17dcca_0800e5ed63c44a23ae9364ca46cddb24.pdf

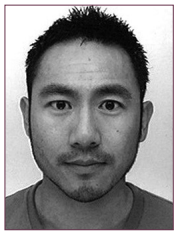
# Image Based Tissue Segmentation:

## Towards the Automation of Mammographic Risk Assessment

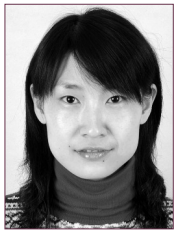


**Reyer Zwiggelaar,**
Professor, Department of Computer Science, Aberystwyth University,
E: rrz@aber.ac.uk

**Harry Strange,**
Postdoctral Research Associate, Aberystwyth University.



**Wenda He,**
Postdoctral Research Associate, Aberystwyth University.



**Zhili Chen,**
Lecturer, Shenyang Jianzhu University.

**Ashwini Kshirsagar,**
Sr Principal Scientist and Manager, Clinical Research, Hologic Inc.

**Erika Denton,**
Honorary Professor of Radiology, University of East Anglia and Norfolk & Norwich University Hospital.

Historically, mammographic risk assessment, i.e. estimating the probability of the development of breast cancer, has been based on an individual's personal and family background. It has been shown that the amount of fibroglandular tissue as well as its distribution of anatomical tissue in mammographic images is strongly correlated with the probability to develop breast cancer. However, manual assessment shows inter- and intra-observer variability and automation of this process has therefore been considered desirable. Such automated methods cover fatty versus dense tissue segmentation, more advanced segmentation approaches and feature space classification. We provide an overview of various approaches to mammographic risk assessment and how this might be used in future computer aided diagnosis (CAD) systems.

### Mammographic risk assessment

Over the past decades, a number of links have been investigated between mammographic risk assessment and patient-specific and environmental aspects, covering family history, diet and genetic markers. Such aspects are currently captured in a number of associated models, e.g. the Gail model [1] and the Tyrer-Cuzick model [2].

However, it should be noted that the above-mentioned mammographic risk models are based on non-image based information and how to integrate image-based information into such risk models is an area of current research [3]. In the late 1960's and mid 1970's, Wolfe [4-6] started to investigate the links between mammographic image information and mammographic risk assessment and found that based on his four risk classes there was a significant difference in risk between the lowest and the highest classes (by a factor of up to twenty in specific studies [5]). Wolfe's classes include aspects of both parenchymal patterns and intensity variations in the mammographic images. This work was followed up by Boyd [7], who established a model based on the percentage dense tissue. Further to this, Tabár and Dean [8] extended the work of both Wolfe and Boyd by describing normal mammographic tissue by four specific building blocks: radiolucent (fatty), homogeneous, nodular and linear tissue, and linked the distribution of these to mammographic risk assessment.

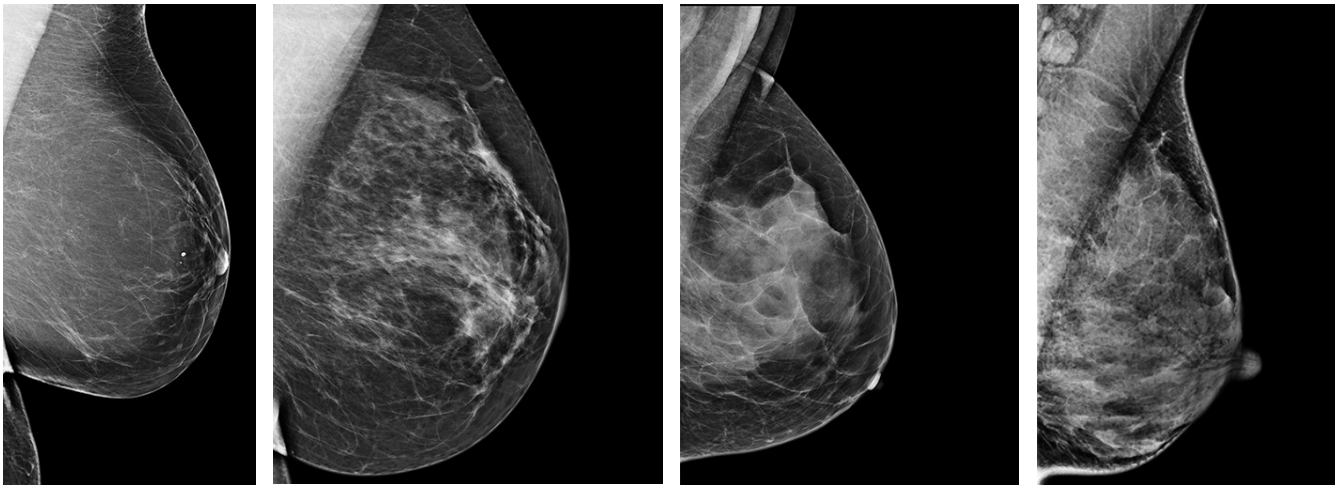
Closely related to Boyd's work, the four Breast Imaging-Reporting and Data System (BIRADS) classes as defined by American College of Radiology BIRADS lexicon are: BIRADS I) the breast is almost entirely fatty, BIRADS II) there are scattered areas of fibroglandular density, BIRADS III) the breasts are heterogeneously dense, which may obscure small masses, and BIRADS IV) the breasts are extremely dense, which lowers the sensitivity of mammography. A set of example mammographic images can be found in Figure 1, which shows BIRADS I to IV cases. Work by Muhimmah et al. [11] has shown that there is clear correlation between the various image based mammographic risk assessment models. Various breast screening and breast cancer detection programmes have adapted mammographic risk assessment [9,10], but it should be noted that none of these currently incorporate the automatic analysis of mammographic image information.

### Dense/fatty tissue

For early research involving automated analysis the methods were closely related to the work of Boyd et al. [7], with a strong emphasis on segmentation and estimation of dense tissue within the breast (the fatty tissue is simply the remaining breast tissue). This work was further developed into "*Cumulus*", which is an interactive software that has been used as a standard within the field [12].

Since the development of Cumulus, there have been a number of approaches that proposed a fully automated method for estimation of dense mammographic tissue. A typical example of this is the recent work by Nickson et al. [13] which provides breast density segmentation based on histogram statistics and boundary gradients information. In contrast, Chen and Zwiggelaar [14] developed an automated density segmentation approach based on fuzzy c-means [15], which incorporates local spatial and intensity information. In both cases, the robustness of the developed approaches was evaluated on large

Figure 1: From left to right mammographic images representing BIRADS I to IV.



datasets. Figure 2 shows how the images shown in Figure 1 are segmented using Chen and Zwiggelaar [14].

Closely related to the described automated work is the development of approaches which incorporated a density-normalised step-wedge into the mammogram capture process [16]. The resulting step-wedge information can be used to estimate the segmentation of differently dense tissue areas, which can in turn be linked with mammographic risk assessment. A slight disadvantage of this approach is that it cannot be used on historical datasets that do not include the step-wedge information.

There have been a number of approaches developed based on the fatty versus dense tissue segmentation work. One of the most successful has been the work by Oliver et al. [17], who provided an initial segmentation of dense and fatty tissue after which they extracted texture and density features from the two regions. The feature space was exploited for the classification of mammographic images into the four BIRADS classes, with correct classification results for the MIAS [18] and DDSM [19] databases of 86% and 77%, respectively. Advanced machine learning techniques were investigated by MacParthalain et al. [20], which improved the classification results to 91% and 89%, respectively.

There has been some work focussing on the links between mammographic risk assessment and volumetric estimation of dense tissue. This takes into account the mammographic projective imaging process. Some of the original work was covered by Highnam and Brady [21]. Additional work was completed by Karssemeijer's research group, which also covered correlation with MRI mammographic data [22].

## Recent approaches

The approaches described in the previous section are based on the distinction between fatty and dense tissue, which only represents part of the clinical descriptions: e.g. both Wolfe and Tabár included parenchymal patterns as part of their classification.

He et al. [23] have used Tabár's work as a foundation to develop mammographic segmentation incorporating anatomical tissue types. Some of this initial work looked at moments as image descriptors, but alternative approaches have also been investigated. The moments-based description provided individual segmentation models for the four (i.e. radiolucent, homogeneous, nodular and linear) tissue types. Such work provides enriched segmentation results as shown in Figure 3. Tabár's original tissue percentages can be directly linked to mammographic

Figure 2: Dense tissue segmentation of the mammograms shown in Figure 1 based on the fuzzy c-means methodology developed by Chen and Zwiggelaar [14].

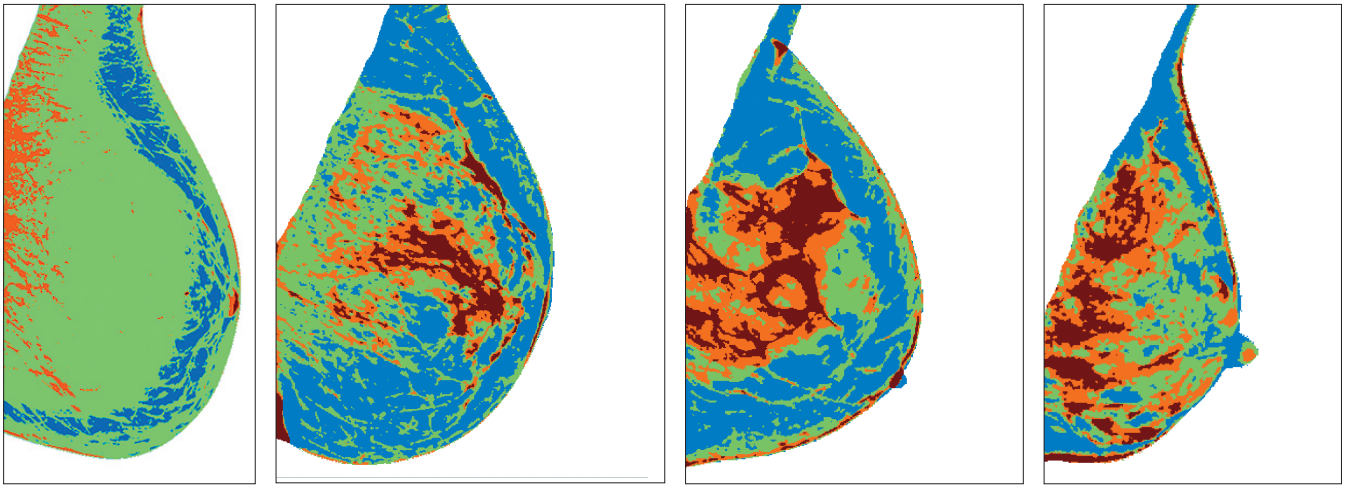
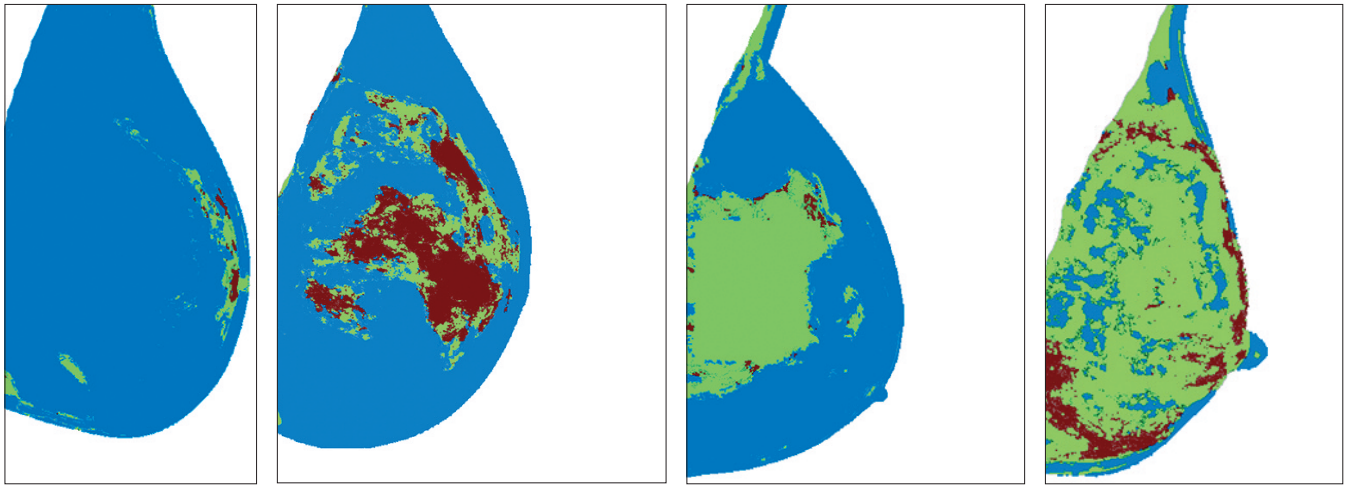


Figure 3: Segmentation of the images shown in Figure 1 using the methodology developed by He et al. [23] which shows radiolucent (blue), nodular (brown) and homogeneous (green) tissue regions.



risk assessment and the most recent results on digital mammographic data show correct classification rates of about 79%. This approach has been evaluated on both digitised and digital mammographic images and has shown robustness with regard to this.

Closely linked to the Tabár tissue type based segmentation developments, Chen et al. [24] have developed a mammographic blob distribution model, which we believe models the homogenous and nodular tissue types. A standard approach to blob detection in images has been adapted to estimate the distribution of blobs at multiple scales and the prior expectation of the two tissue types has been taken into account. In Figure 4 a blob representation of a set of example mammograms is shown, which indicates how the multi-scale blob distribution changes with the BIRADS density classification. This resulted in a number of metrics, which were linked to mammographic risk assessment and achieved classification accuracies close to 80% for the MIAS database.

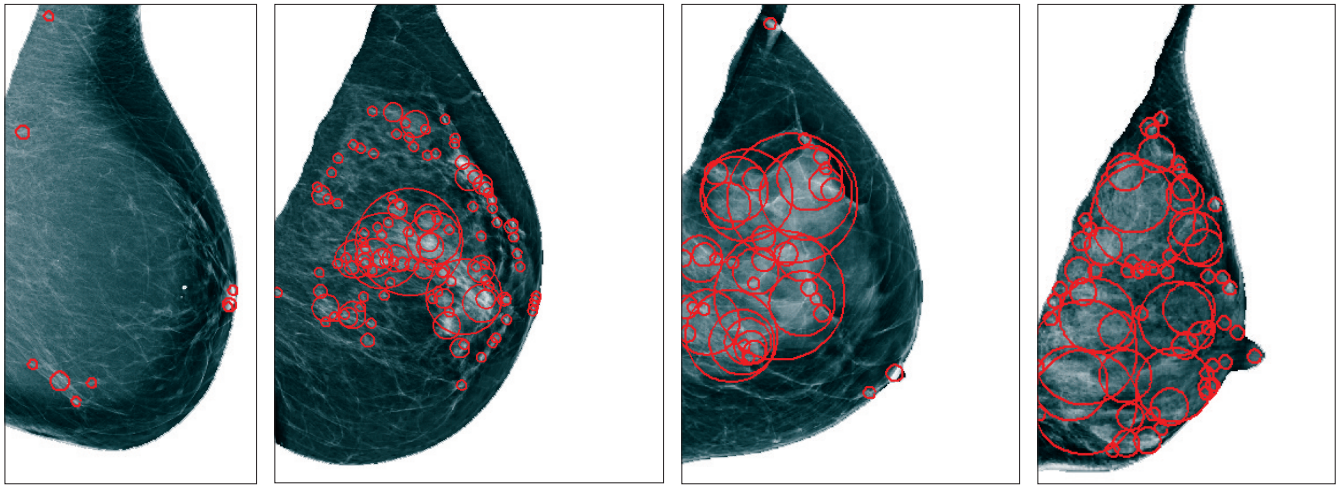
Over recent years we have also started to incorporate both topology and manifold learning techniques into our computer vision approaches. Such techniques provide low dimensional models of data in high dimensional feature space and can be used for dimensionality reduction and noise suppression. We have used this to obtain improved segmentation of the dense regions in mammograms [25] with initial results on a limited dataset showing significant improvements. An alternative segmentation technique was developed by Chen et al. [24], which was based on a topographic map of the whole breast, representing both topographic and geometrical structures. Their initial results indicate the potential for advanced techniques to make a contribution to mammographic risk assessment.

## Future directions

The main aim of the development of mammographic risk assessment techniques is to identify women at high risk of getting breast cancer in future and then triage them into optimal paths of screening, diagnosis and treatment paradigms. It is hoped that novel risk assessment techniques will eventually integrate into commercial CAD systems. There are two potential aspects for which the risk assessment based on density or tissue segmentation approaches can be used: 1) the identification of high-risk cases, which could receive additional attention, different subsequent imaging such as ultrasound or breast MR and/or could be invited more often for screening, and 2) as input information for fully automated computer detection algorithms. Both these avenues could increase the probability of detecting breast cancer at an early stage.

It should be noted that there are a

Figure 4. Blob representation of the images shown in Figure 1 using the methodology developed by Chen et al. [24].



number of commercial systems available, which are currently aimed at estimating volumetric breast density and/or area breast density for Full-Field Digital Mammography (FFDM) images. Such systems are used to assist radiologists in the assessment of breast tissue composition and provides a density score, which can be linked to BIRADS breast composition categories. These systems include *Quantra* (Hologic Inc.), *Volpara* (Volpara Solutions), and *MicroDose SI* (Philips Healthcare).

The systems described above are aimed at helping mammography practices achieve a reader independent objective breast density assessment. After the adoption of tomosynthesis images, it is expected that conventional FFDM images will phase out and, there will be a need to translate the currently developed approaches of breast density estimation using FFDM to the new modality of tomosynthesis. This would possibly provide a close to volumetric density/tissue segmentation, which may lead to a more reliable mammographic risk assessment.

It should also be noted that another aspect for further development of the Tabár-based work is the temporal analysis of not just changes in density, but also in parenchymal patterns. ●

REFERENCES


1. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, Mulvihill JJ. *Projecting individualized probabilities of developing breast cancer for white females who are being examined annually.* J Natl Cancer Inst 1989;81(24):1879–86.
2. Tyrer J, Duffy SW, Cuzick J. *A breast cancer prediction model incorporating familial and personal risk factors.* Stat Med 2004;23,1111-30.
3. Eriksson L, Hall P, Czene K, Dos Santos SI, McCormack V, Bergh J, Bjohle J, Ploner A. *Mammographic density and molecular subtypes of breast cancer.* Br J Cancer 2012;107,18–23.
4. Wolfe JN. *A study of breast parenchyma by mammography in the normal woman and those with benign and malignant disease.* Radiology 1967;9,201-205.
5. Wolfe JN. *Breast patterns as an index of risk for developing breast cancer.* American Journal of Roentgenology 1976a;126,1130-7.
6. Wolfe JN. *Risk for breast cancer development determined by mammographic parenchymal pattern.* Cancer 1976b;37(5), 2486-92.
7. Boyd NF, O'Sullivan B, Campbell JE, Fishell E, Simor I, Cooe G, Germanson T. *Mammographic signs as risk factor for breast cancer.* Br J Cancer 1982;45, 185-93.
8. Tabár L, Dean PB. *Mammographic parenchymal patterns and risk of breast cancer?* J Am Med Assoc. 1982;247,185-98.
9. BIRADS (1995) American College of Radiology. Breast Imaging Reporting and Data System (BI-RADS) 2nd edition. Reston: American College of Radiology.
10. Roche NA, Given-Wilson RM, Thomas VA, Sacks NP. *Assessment of a scoring system for breast imaging.* Br J Surgery 1998;85,669-72.
11. Muhimmah I, Oliver A, Denton ERE, Pont J, Pérez E, Zwiggelaar R. *Comparison between Wolfe, Boyd, BI-RADS and Tabár based mammographic risk assessment.* LNCS 2006;4046, 407-415.
12. Yaffe MJ. *Measurement of mammographic density.* Breast Cancer Research 2008;10(209).
13. Nickson C, Arzhaeva Y, Aitken Z, Elgindy T, Buckley M, Li M, English DR, Kavanagh AM. *AutoDensity: an automated method to measure mammographic breast density that predicts breast cancer risk and screening outcomes.* Breast Cancer Research 2013;15,R80.
14. Chen Z, Zwiggelaar R. *A modified fuzzy c-means algorithm for breast tissue density segmentation in mammograms.* 10th IEEE International Conference on Information Technology and Applications in Biomedicine 2010.
15. Dunn JC. *A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters.* Journal of Cybernetics 1973;3,32-57.
16. Diffey J, Hufton A, Astley SM. *A new step-wedge for the volumetric measurement of mammographic density.* Lecture Notes in Computer Science 2006;4046,1-9.
17. Oliver A, Freixenet J, Marti R, Pont, J, Pérez E, Denton ERE, Zwiggelaar R. *A novel breast tissue density classification methodology.* IEEE Transactions on Information Technology in Biomedicine 2008;12(1),55-65.
18. Suckling J, Partner J, Dance DR, Astley SM, Hutt I, Boggis CRM, Ricketts I, Stamatakis E, Cerneaz N, Kok SL, Taylor P, Betal D, Savage J. *The Mammographic Image Analysis Society digital mammogram database.* International Workshop on Digital Mammography 1994;211-21.
19. Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer PJ. *The Digital Database for Screening Mammography.* Proceedings of the International Workshop on Digital Mammography 2000;212–8.
20. MacParthalain N, Jensen R, Shen Q, Zwiggelaar R. *Fuzzy-rough approaches for mammographic risk analysis.* Intell Data Anal 2010;14 (2),225-44.
21. Hartman K, Highnam R, Warren R, Jackson V. *Volumetric assessment of breast tissue composition from FFDM images.* LNCS 2008;5116,33-9.
22. van Engeland S, Snoeren PR, Huisman H, Boetes C, Karssemeijer N. *Volumetric breast density estimation from full-field digital mammograms.* IEEE Transactions on Medical Imaging 2006;25(3),273-82.
23. He W, Denton ERE, Stafford K, Zwiggelaar R. *Mammographic image segmentation and risk classification based on mammographic parenchymal patterns and geometric moments.* Biomedical Signal Processing and Control 2011;6(3),321-9.
24. Chen Z, Denton ERE, Zwiggelaar R. *Topographic representation based breast density segmentation for mammographic risk assessment.* International Conference on Image Processing 2012;1993-6.
25. Strange H, Denton ERE, Kibiro M, Zwiggelaar R. *Manifold learning for density segmentation in high risk mammograms.* Lecture Notes in Computer Science 2013;7887,245-52.
26. Chen Z, Wang L, Denton E, Zwiggelaar R. *A multiscale blob representation of mammographic parenchymal patterns and mammographic risk assessment.* Lecture Notes in Computer Science 2013;8048,346-53.