# IWDM 2000
# 5th International Workshop on Digital Mammography

**MARTIN J. YAFFE**
EDITOR

**Proceedings of the Workshop**
**June 11-14, 2000**
**Toronto, Canada**

**Medical Physics Publishing**
**Madison, Wisconsin**

# Development and Assessment of Protocols for Efficient Utilization of Large-Scale Digital Mammography Databases

**MELANIE A. SUTTON**
**ADRIANA MARIN**
**SHERI SENTELLE**
*Robotics and Image Analysis Laboratory*
*Department of Computer Science, University of West Florida*
*Pensacola, Florida*

**SUZANNE WOOLDRIDGE**
**PATRICIA GRIFFITH**
*Ann L. Baroco Center for Women's Health, Sacred Heart Hospital*
*Pensacola, Florida*

**CAROL DELL**
*West Florida Regional Medical Center, Pensacola, Florida*

## INTRODUCTION

As the need for a second tireless mammogram reader has increased, research has turned to the development of computer-aided diagnosis (CAD) systems. We specifically investigated one aspect of these CAD systems, the need for sampling techniques that can be used to enhance and analyze digital mammograms with less storage and time requirements. We limited our initial studies of this approach to two commonly cited segmentation techniques: the supervised k-nearest neighbor (k-nn) algorithm (Devijver and Kittler 1982) and the unsupervised Fuzzy C-Means (FCM) algorithm (Bezdek et al. 1999). Shankar and Pal (1994) introduced the methodology for the Fast Fuzzy C-Means (Fast FCM) algorithm studied here, which makes use of representative data sets ($\mathbf{X}_{\text{sample}}$) extracted from the full image ($\mathbf{X}_{\text{full}}$) to derive prototypes which can then be used to cluster the complete data set. To assess the utility of this approach in the mammography domain, we were fortunate to be able to work with two radiologists who contributed images with annotated ground truth information to the *Digital Database for Screening Mammography* (DDSM) (Heath et al. 1998). Figure 1 presents the overall flow of control for the prototype mammographic interpretation system described in this paper.

## STATISTICAL SAMPLING

The chi-square test is a statistical test commonly used to compare observed data with data that would be expected if assuming some specific hypothesis (Bain and Engelhardt 1992). For the work described here, to derive a sample from a mammographic image, the chi-square test is used in two ways to determine the "goodness of fit" between the observed data (in the sampled image) and the expected data (in the full image):

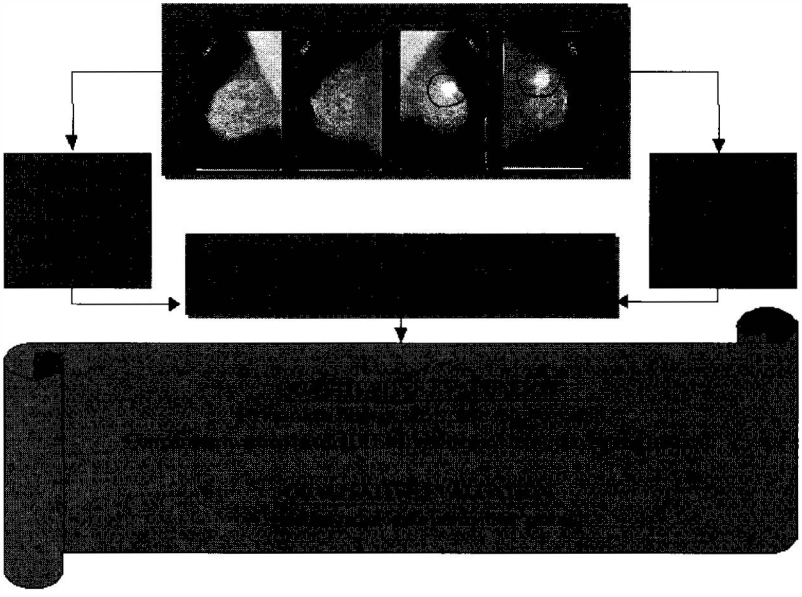

**Figure 1.** Overview of mammographic interpretation system processing steps.

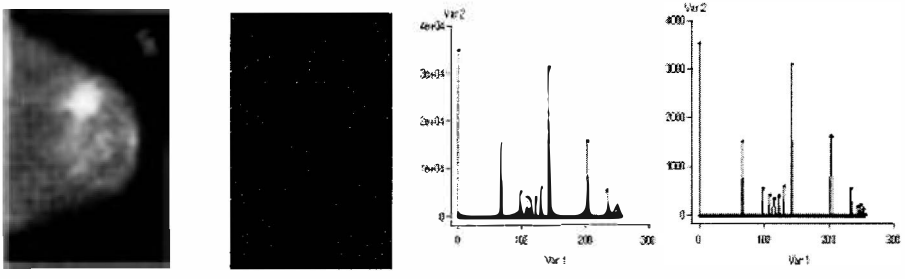

**Figure 2.** Outputs utilized to derive a statistically representative sample from a DDSM image. (a) Original case0028_LCC image representing $\mathbf{X}_{\mathbf{full}}$. (b) Binary image mask representing $\mathbf{X}_{\mathbf{sample}}$ pixel locations (white areas indicate locations of 10% of $\mathbf{X}_{\mathbf{full}}$). (c) Histogram based on $\mathbf{X}_{\mathbf{full}}$. (d) Histogram based on $\mathbf{X}_{\mathbf{sample}}$ (Var 1 = Intensity Value; Var 2 = Number of Pixels).

1. To test the randomly selected row and column locations to ensure that the image is uniformly spatially sampled so that statistically significant image areas are not omitted; and
2. To test the quality of the extracted features underlying the image locations from usage 1 above, to ensure the sample is statistically significant in terms of the diversity of the data in the sample.

Figure 2 provides an example of outputs from these steps for an image from DDSM, utilizing intensity information derived from pixel locations.

**Table 1. Performance Metrics Used for Analysis of Results**

| Metric | Description | Used in |
|---|---|---|
| Amount of memory | # of bytes for full and sampled images | $\mathbf{X}_{full}$ to $\mathbf{X}_{sample}$ comparisons for k-nn and FCM |
| Time | # of seconds to run the specified algorithm | $\mathbf{X}_{full}$ to $\mathbf{X}_{sample}$ comparisons for k-nn and FCM |
| *compareTemplates* - percentage overlap, centroid localization | DDSM tool utilizing 4-connected region-based TP/FP numbers | $\mathbf{X}_{full}$ to $\mathbf{X}_{gt}$ comparisons for k-nn |
| Pixel-based metrics | Counts of TP/FP/FN/TN locations | $\mathbf{X}_{full}$ to $\mathbf{X}_{gt}$ comparisons for k-nn and FCM |
| % Difference | Percentage difference in labels assigned by FCM and FFCM | $\mathbf{X}_{full}$ to $\mathbf{X}_{sample}$ comparisons for FCM |
| Acceleration factor | Ratio of CPU time taken by FCM and FFCM | $\mathbf{X}_{full}$ to $\mathbf{X}_{sample}$ comparisons for FCM |
| Storage improvement factor | Ratio of memory used for FCM and FFCM | $\mathbf{X}_{full}$ to $\mathbf{X}_{sample}$ comparisons for FCM |

The histograms of figure 2(c) and 2(d) are visually similar and were confirmed to be statistically similar at the 95% confidence level with the chi-square test.

## METHODOLOGY FOR ALGORITHMIC ASSESSMENT

As described in figure 1, the evaluation methodology developed to determine the success or failure of the enhancement algorithms with or without sampling places emphasis on both subjective criteria (to determine the utility to practicing radiologists) and on objective, quantitative criteria (using database ground truth information). For DDSM images, the assessment process begins with the use of the *createTemplate* tool to derive a binary ground truth image (i.e., 0 for normal areas, 1 for abnormal areas).[1] Various methods for quantitative analysis are then possible, using metrics available at the DDSM site with the *compareTemplates* tool (based on percentage overlap or centroid localization of algorithmic and ground truth regions) and/or traditional pixel-based analysis. The complete set of performance metrics utilized in our work is described in table 1.

## EXPERIMENTAL RESULTS

The metrics shown in table 1 were used to determine whether segmentation results, generated faster using sampled data, were "comparable" to the results obtained on the full data set, without loss of medically relevant information. To test this objective, subsets of DDSM images with and without abnormalities were obtained and processed. Image case0028_LCC was used during training to

[1]Tools for creating ground truth templates and deriving quantitative comparisons can be obtained from the DDSM site (http://marathon.csee.usf.edu/Mammography/Database.html).

**Table 2. Summary of Image Types and Uses for DDSM Cases 0028, 0037, and 0051**

| Case # | Application | View | Type of Abnormality |
|---|---|---|---|
| case0028 | • Training (parameter set development) | LCC | Spiculated mass |
| | • Testing (k-nn and FCM) | LCC | Spiculated mass |
| | • Testing (k-nn and FCM) | LMLO | Spiculated mass |
| | • Testing (k-nn and FCM) | RMLO, RCC | Normal |
| case0037 | • Testing (FCM) | RMLO, RCC | Normal |
| | • Testing (FCM) | LMLO, LCC | Spiculated mass |
| case0051 | • Testing (FCM) | RMLO, RCC | Microlobulated mass |
| | • Testing (FCM) | LMLO, LCC | Normal |

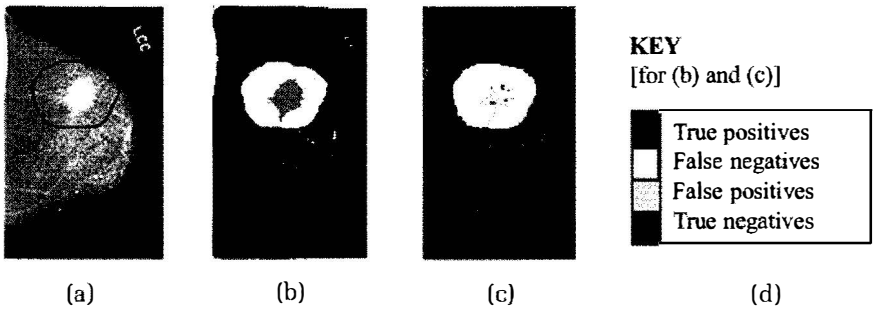


**Figure 3.** Example input and output images for segmentation based on k-nn algorithm. (a) Original case0028_LCC image. (b) Segmentation results based on k = 5 for $\mathbf{X}_{\text{full}}$. (c) Segmentation results based on k = 5 for $\mathbf{X}_{\text{sample}}$, representing 10% of $\mathbf{X}_{\text{full}}$ (d) color key.

determine the range of operating points for clinically useful parameter sets. The remaining images were utilized as described in table 2.

### Analysis of the k-nn Segmentation Algorithm

The k-nn algorithm is a supervised technique that uses physical labels of tissue classes selected by humans prior to segmentation (Devijver and Kittler 1982). For this experiment, the training data were selected by hand from 900 locations in image case0028_LCC to cover three classes (tumor tissue, normal breast tissue, and background). The distance metric utilized in all experiments was the Euclidean norm. Figure 3 shows the results of k-nn processing on the training image to demonstrate the utility of this approach. Similar to commercial mammographic interpretation systems such as the R2 ImageChecker system (R2 Technology 1998), our ultimate objective is to provide a small set of ranked prompts on the input mammogram by additional analysis of the outputs shown in figure 3(b) and 3(c).

**Table 3. Performance Results for Case0028_LCC ($X_{full}$ and $X_{sample}$) Compared to Ground Truth**

| | Pixel-based Metrics | | | | Time (seconds) | | Memory (bytes) | |
|---|---|---|---|---|---|---|---|---|
| | TP | | FP | | | | | |
| k | $X_{full}$ | $X_{sample}$ | $X_{full}$ | $X_{sample}$ | $X_{full}$ | $X_{sample}$ | $X_{full}$ | $X_{sample}$ |
| 5 | 2875 | 561 | 2147 | 235 | 53 | 5 | 529052 | 52920 |
| 11 | 2875 | 343 | 2147 | 44 | 61 | 5 | 529076 | 52944 |
| 27 | 2875 | 400 | 2147 | 154 | 89 | 6 | 529140 | 53008 |
| 100 | 2896 | 300 | 2067 | 0 | 251 | 9 | 529432 | 52300 |
| 200 | 3271 | 300 | 2657 | 0 | 749 | 15 | 529832 | 53700 |

**Table 4. Performance Results for Case0028_LCC ($X_{full}$) Utilizing Region-Based Ground Truth Comparisons**

| | DDSM *compareTemplates* Percentage Overlap Metric ($t_{overlap}$ = 0.10) | | DDSM *compareTemplates* Centroid Localization Metric | |
|---|---|---|---|---|
| k | TP | FP | TP | FP |
| 5,11,27 | 1 | 72 | 1 | 53 |
| 100 | 1 | 63 | 1 | 39 |
| 200 | 1 | 119 | 1 | 85 |

As summarized in tables 3 and 4, sampling generates a significantly smaller set of candidate pixels or regions in times that are more realistic for clinical settings.

### Analysis of the FCM Segmentation Algorithm

For comparison to an unsupervised approach, the entire set of images from case0028, case0037, and case0051 was provided as input to the FCM algorithm (Bezdek et al. 1999) to determine how the algorithm behaved on images with and without abnormalities. The initial set of parameters utilized in this experiment was derived from feedback from radiologists who reviewed image outputs based only on processing of image case0028_LCC. Based on their feedback, the parameter settings for FCM for the remainder of the analysis were established as: c = 2, 5, or 9 for the number of classes, initialization using the diagonal of the hyperbox of the feature space, a termination criteria of 50 iterations or prototype difference of 0.5, and a weighting exponent of 2.0. All distance calculations involved the Euclidean norm.

Figure 4 shows representative input and output images from this algorithm for images with and without abnormalities. For the purposes of our interpretation assistance system, our design allows the radiologist to view sequences of

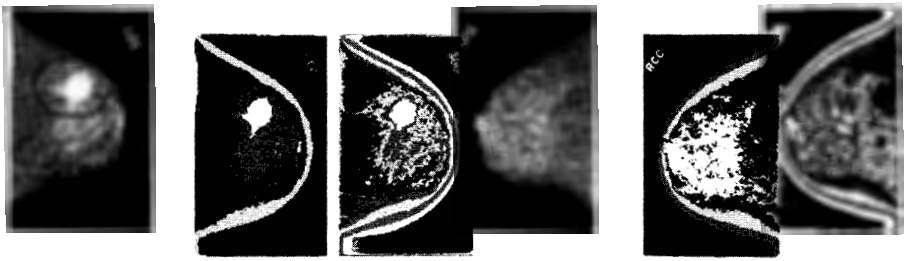


**Figure 4.** Example input and output images for segmentation based on the FCM algorithm. (a) Original case0028_LCC image. (b) Segmentation results based on c = 5 for $\mathbf{X}_{\text{sample}}$. (c) Segmentation results based on c = 9 for $\mathbf{X}_{\text{sample}}$. (d) Original case0028_RCC image. (e) Segmentation results based on c = 5 for $\mathbf{X}_{\text{sample}}$. (f) Segmentation results based on c = 9 for $\mathbf{X}_{\text{sample}}$.

FCM outputs such as those shown in figure 4 to assess the impact of a suspected abnormal area on surrounding tissues. The output produced from FCM during sampling is a $1 \times |\mathbf{X}_{\text{sample}}|$ text file that only represents the class labels of the sampled points in the image. To generate the images shown in figure 4, the remaining unsampled data points were assigned class labels by using the resulting prototypes from FCM to determine their membership values. Once this step was completed, the results were combined to form a single PGM image containing class labels for all points in the original image.

As indicated in table 5, sampling results in significant improvements in time and storage for this algorithm. The acceleration factors indicate speed improvements ranging from 17% to 426% faster. Furthermore, the labels assigned by FCM with and without sampling after hardening are similar, with average differences in the range of 6% to 12%. A quantitative evaluation of the final prototype values for $\mathbf{X}_{\text{sample}}$ and $\mathbf{X}_{\text{full}}$ indicated differences of around 5 intensity values.

## DISCUSSION AND FUTURE INVESTIGATIONS

This paper explored aspects of digital mammography for the purpose of designing and incorporating modules into a prototype mammographic enhancement system that might be useful in clinical environments. A technique for sampling an image to obtain a statistically representative subset of pixels (based on the chi-square statistical test) was presented. This technique was useful for minimizing the impact of large image size on storage and time requirements. Furthermore, a visual comparison between the outputs of the algorithms on full images and the sampled results created with just 10% to 15% of the original data indicates that the sampled image results are comparable in terms of retaining medically relevant information for practicing clinicians.

In terms of improving system intelligence, we are in the process of incorporating the extraction of additional edge-based features studied in previous work (Sutton and Bezdek 1998, Bezdek and Sutton 1999). We will continue to work closely with clinicians to incorporate this information and to determine

**Table 5. Average Results for c = 2, 5 and 9 for FCM Processing With and Without Sampling**

| Case # | % of Data Used | % Difference ± Std. Dev. | Acceleration Factor ± Std. Dev. | Storage Improvement Factor ± Std. Dev. |
|---|---|---|---|---|
| C0028_LCC | 10 | 6.20 ± 5.40 | 1.17 ± 0.47 | 10.0 ± 5.96 E-6 |
| C0028_LMLO | 10 | 6.31 ± 6.27 | 1.53 ± 0.40 | 10.0 ± 5.77 E-6 |
| C0028_RCC | 15 | 8.11 ± 7.71 | 2.44 ± 1.65 | 6.67 ± 2.36 E-6 |
| C0028_RMLO | 10 | 7.53 ± 7.31 | 2.06 ± 0.54 | 10.0 ± 5.58 E-6 |
| C0037_LCC | 15 | 8.20 ± 10.51 | 3.23 ± 1.27 | 6.67 ± 1.83 E-6 |
| C0037_LMLO | 10 | 11.71 ± 13.55 | 4.26 ± 2.04 | 10.0 ± 5.83 E-6 |
| C0037_RCC | 10 | 8.07 ± 7.34 | 2.25 ± 1.25 | 10.0 ± 5.67 E-6 |
| C0037_RMLO | 10 | 7.53 ± 7.31 | 3.96 ± 1.85 | 10.0 ± 5.53 E-6 |
| C0051_LCC | 15 | 7.01 ± 5.97 | 2.78 ± 0.83 | 6.67 ± 1.17 E-5 |
| C0051_LMLO | 10 | 7.07 ± 5.19 | 4.06 ± 1.92 | 10.0 ± 5.89 E-6 |
| C0051_RCC | 10 | 12.04 ± 12.24 | 4.10 ± 2.06 | 10.0 ± 5.73 E-6 |
| C0051_RMLO | 10 | 5.17 ± 4.26 | 4.08 ± 2.41 | 10.0 ± 5.23 E-6 |

other relevant features which improve the sensitivity and specificity of the overall system.

There are many additional avenues to be explored to ensure the development of a robust, clinically useful system for detecting all the warning signs of breast cancer (masses, microcalcifications, architectural distortions, etc.). While sampling helped to improve the speed of algorithmic processing for the segmentation techniques studied here, additional optimization methods need to be investigated, as radiologists often can assess mammograms in a matter of seconds (not minutes). Additional research is also needed to address questions of whether training data developed from images derived from one scanner can be applied to images derived from other scanners. Finally, to compare our results to commercially available mammogram prompting systems, we are obtaining R2 ImageChecker system outputs for the images described in this paper (using the original films supplied to DDSM). As the fight against breast cancer continues, and as more and more researchers begin using the completed DDSM data set, we look forward to the challenge of comparing results from our final system with those of other researchers in the international community.

## ACKNOWLEDGMENTS

This research was funded by a Whitaker Foundation Biomedical Engineering Research Grant.

We would like to thank Chris Sentelle for programming assistance during the development of the code for the FCM algorithm. Researchers interested in

obtaining a copy of the code for their own research may download it from our FTP site: http://www.cs.uwf.edu/~msutton/lab.

## REFERENCES

Bain, L. J., and M. Engelhardt. *Introduction to Probability and Mathematical Statistics*, 2nd Edition. Belmont, CA: Duxbury Press, 1992.

Bezdek, J. C., and M. A. Sutton. "Image processing in medicine." in *Practical Applications of Fuzzy Technologies*. H.-J. Zimmermann (ed.). Norwell, MA: Kluwer Academic Publishers, pp. 363–418, 1999.

Bezdek, J. C., J. Keller, R. Krisnapuram, and N. R. Pal. *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. Boston, MA: Kluwer Academic Publishers, 1999.

Devijver, P. A., and J. Kittler. *Pattern Recognition: A Statistical Approach*. Engelwood Cliffs, NJ: Prentice-Hall International, Inc., 1982.

Heath, M., K. W. Bowyer, D. Kopans, P. Kegelmeyer, R. Moore, K. Chang, and S. Munishkumaran. "Current Status of the Digital Database for Screening Mammography" in *Digital Mammography*. N. Karssemeijer, M. Thijssen, J. Hendriks, and L. van Erning (eds.), Proceedings of the 4th International Workshop on Digital Mammography, Nijmegen, The Netherlands, June 1998. Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 457–460, 1998. (Electronic contact information: http://marathon.csee.usf.edu/Mammography/Database.html).

R2 Technology, Inc. R2 ImageChecker M1000 System. Los Altos, CA, 1998. (http://www.R2Tech.com).

Shankar, B.U., and N. R. Pal. "FFCM: An Effective Approach for Large Data Sets" in *Proceedings of the Third International Conference On Fuzzy Logic, Neural Nets and Soft Computing*, IIZUKA, pp. 331–332, 1998.

Sutton, M. A., and J. Bezdek. "Enhancement and Analysis of Digital Mammograms Using Fuzzy Models" in *Proceedings of the 26th Applied Imagery and Pattern Recognition (AIPR) Workshop: Exploiting New Image Sources and Sensors (SPIE Vol. 3240)*. J.M. Selander (ed.). Bellingham, WA: SPIE Press, pp. 179–190, 1998.