

Computerized Diagnosis of Breast Fine-Needle Aspirates

William H. Wolberg, M.D.,* W. Nick Street, Ph.D.,*[†]
and Olvi L. Mangasarian, Ph.D.[†]

Departments of Surgery and Computer Sciences Department,[†]
University of Wisconsin, Madison*

■ **Abstract:** The goal of this work is to determine the accuracy of computer-based image analysis in diagnosing breast fine-needle aspirates (FNA). On 192 FNAs, the computer-based diagnostic accuracy was 97.9%. This is consistent with the 97.5% accuracy projected by machine learning methods during the initial training in 1994 with 569 FNAs. One of the attributes of this system is the rendering of a value that estimates the probability of malignancy. We consider estimated probability of malignancy values between 0.30 and 0.70 to be equivocal. Eleven of our samples (5.7%) fell into this equivocal category. All computer misclassified FNAs were correctly diagnosed visually based on contextual features. We propose this computer-based system as a diagnostic adjunct rather than as a stand-alone system. ■

Key Words: Breast cancer, diagnosis, computer analysis, machine learning, fine-needle aspirates

The goal of this work is to determine the accuracy of computer-based image analysis in diagnosing breast fine-needle aspirates (FNA). This paper describes the prospective accuracy of an interactive computer-based diagnostic system that we call Xcyt. Using a regular optical microscope on which a small videocamera is mounted, an operator selects the FNA field to be analyzed. The image is then displayed on a Unix-based workstation and representative nuclei are outlined by the user with a mouse. The program then adjusts the

outline to conform to the exact nuclear border. The system calculates 10 nuclear size, shape, and textural features and reports their mean, standard errors, and worst values. Using machine learning methods, an algorithm was developed on the basis of an initial series of 569 patients (357 benign and 212 malignant) with established diagnoses. Tenfold cross-validation (1) of these data led us to believe that 97.5% accuracy could be achieved. In this article, we report the prospective computer diagnosis of 192 consecutive FNAs.

MATERIALS AND METHODS

A consecutive series of 569 patients (212 cancer, 357 benign) provided the data to develop (train) the algorithm. Since then, an additional 194 consecutive, new patients were aspirated. Two samples did not contain sufficient cells for analysis. The 192 satisfactory FNAs were used to test the system. All 63 cancers and 33 of the benign masses were histologically confirmed. The 96 remaining benign masses were followed for a year without change in size or character. The TNM classification of the FNA diagnosed cancers is given in Table 1.

Fine-Needle Aspiration

A small drop of viscous fluid was aspirated from breast masses by making multiple passes with a 23-gauge needle as negative pressure was applied to an attached syringe. The aspirated material was expressed onto a silane-coated glass slide. A similar slide was placed face-down on the aspirate, and the aspirate was spread as the slides were separated with a horizontal motion. Preparations were immediately fixed in 95% ethanol and examined after they were stained with hematoxylin and eosin. Most of the samples were obtained

Address correspondence and reprint requests to: William H. Wolberg, M.D., Department of Surgery, University of Wisconsin, 600 Highland Avenue, Madison, WI 53792, U.S.A.

Table 1. TNM Tabulation of 61 of the FNA-Diagnosed Cancers (In two cases histologic confirmation was obtained by autopsy)

	<i>n</i> = 0	<i>n</i> = 1	<i>n</i> = x
T =			
1 IS	1	0	2
1a	7	8	5
1b	0	0	1
2a	8	15*	2*
3 IS	0	0	1
3a	2	4*	1
3b	0	1	0
3c	1	0	0
4b	0	1	0
4d	0	1	0

*One patient in each of these groups had distant metastases at the time of initial diagnosis.

from palpable masses although several recent ones were obtained under ultrasound guidance. Only solid masses that yielded epithelial cells were computer-analyzed.

Image Preparation

The area on the aspirate slides to be analyzed was visually selected by an operator for minimal nuclear overlap. The image for digital analysis was generated by a JVC TK-1070U color videocamera mounted atop an Olympus microscope and the image was projected into the camera with a 63× objective and a 2.5× ocular. The image was captured by a ComputerEyes/RT color framegrabber board (Digital Vision, Inc., Dedham MA 02026) as a 640 × 400 Targa file. An 8-bit-per-pixel gray-scale image is used for the image analysis, since we are not interested in the color of the nuclei.

Nuclear Features

The first step in successfully analyzing the digital image is to specify the exact location of each cell nucleus. A graphical computer program called Xcyt was developed that allows the user to input the approximate location of enough nuclei (10–20) to provide a representative sample. A mouse is used to trace a rough outline of cell nuclei on the computer monitor. From this rough outline, the actual boundary of the cell nucleus is located by an adaptive spline technique (2). The mathematical aspects of these calculations are described elsewhere (3).

Once the nuclei to be analyzed have been identified by the operator and have been enclosed by the computer-generated splines, the computer calculates 10 nuclear features for each nucleus (3). These features are modeled such that higher values are typically associated with malignancy. Nuclear size is expressed by the radius

and area features. Nuclear shape is expressed by smoothness, concavity, compactness, concave points, symmetry, and fractal dimension features. Both size and shape are expressed by the perimeter feature. Nuclear texture is measured by finding the variance of the gray-scale intensities in the component pixels. Features were verified using idealized phantom cells (4). The mean value, worst (mean of the three largest values), and standard error of each feature are computed for each image, resulting in a total of 30 features.

Data Analysis

The mathematical programming on which the data analysis is based has been described in detail elsewhere (5). Briefly, image processing produces a database consisting of one 30-dimensional point for each sample. We then wish to construct a surface that separates the benign points from the malignant points in this 30-dimensional space. The classification procedure is known as MSM-Tree (MSM-T) (6,7). This method uses linear programming to iteratively place a series of separating planes in the feature space of the samples. If the benign and malignant sets can be separated by a single plane, the first plane will be so placed between them. If the sets are not linearly separable, MSM-T constructs a plane that minimizes an average distance of misclassified points. Depending on the separation accuracy attained, the procedure is recursively repeated on the two regions generated by each plane until satisfactory separation is achieved, that is, each of the final regions contains mostly points of one category. The classifier thus obtained is then used as a decision tree to categorize new cases. MSM-T is similar to other decision-tree methods such as CART (8) and C4.5 (9) but has been shown to be faster and more accurate on several real-world data sets (6).

In many instances, simpler classifiers perform better on new data than do more complex ones. Therefore, we minimize not only the number of separating planes but also the number of features used in constructing the planes. The three diagnostic features that, when used together, gave the best separation were worst area, worst smoothness, and mean texture (3). These three features and a single separating plane were used to accomplish diagnostic separation. Using cross-validation (1) the diagnostic accuracy of the resulting classifier was estimated to be 97.5%.

Our computerized system gives a benign or malignant diagnosis together with a probability of malignancy determined by the distance the new point lies

Table 2. Computer-Generated Estimated Probability for 192 Prospectively Obtained Breast FNAs and their Eventual Diagnosis

Xcyt Estimated Probability of Malignancy	Final Diagnosis	
	Benign	Malignant
0.00 to 0.09	98*	0
0.10 to 0.19	14	1 [†]
0.20 to 0.29	5	0
0.30 to 0.39	8	0
0.40 to 0.49	1	0
0.50 to 0.59	1	0
0.60 to 0.69	1	0
0.70 to 0.79	0	2
0.80 to 0.89	0	1
0.90 to 0.1.00	1 [‡]	59

*One sample from multiple papillomatosis was incorrectly visually diagnosed as malignant.

[†]Lobular neoplasia correctly diagnosed visually.

[‡]Fibroadenoma.

from the separating plane (4). Should this probability fall between 30% and 70%, the sample is considered "suspicious," in the same way a visually graded FNA may be termed suspicious.

RESULTS

Table 2 lists the Xcyt-generated estimated probability of malignancy for the 192 consecutive FNAs from which an adequate diagnostic sample was obtained. One hundred twenty-nine proved to be benign (33 by biopsy and 96 by follow-up) and 63 proved to be malignant.

Xcyt prospectively diagnosed breast FNAs with 97.9% accuracy.

DISCUSSION

Adequate diagnostic FNAs were obtained from 192 of 194 consecutively aspirated masses. The prospective 97.9% accuracy of computer-based diagnosis is consistent with the 97.5% accuracy projected by machine-learning methods during Xcyt's initial training in 1994 with 569 FNAs. In addition to validating the accuracy of Xcyt, these results also confirm the use of cross-validation of retrospective data as a means for estimating prospective accuracy.

Correct visual diagnoses can easily be made for FNA samples at the extremes of malignancy and benignity. Diagnostic difficulties arise for equivocal samples. One of the attributes of Xcyt is the rendering of a value that estimates the probability of malignancy. Currently, we consider Xcyt-estimated probability of malignancy values between 0.30 and 0.70 to be equivocal. Eleven of our samples (5.7%) fell into this equivocal category. All

were benign and included two that fell on the malignant side of the separating plane. The other two misclassifications were a lobular neoplasia (lobular carcinoma in situ) that was clearly misclassified as benign and a fibroadenoma that was clearly misclassified as malignant. Xcyt does not address the problem of inadequate samples that constitute a small percentage in this (4% of samples) and in other series (10–13).

The estimated probability of malignancy rendered by Xcyt is clinically useful by objectively quantitating the uncertainty inherent in classifying some FNAs. Visually, such FNAs are reported as "suspicious." This estimated probability of malignancy as rendered by Xcyt provides objectivity for discussion the need for following up on the FNA results, be it observation, excisional biopsy, or definitive surgery.

We propose Xcyt as a diagnostic adjunct rather than as a stand-alone system. All computer misclassified FNAs were correctly diagnosed visually. Contrariwise, one FNA obtained from a case of multiple polyposis was incorrectly diagnosed visually, but was correctly diagnosed by Xcyt. Although Xcyt accurately assesses single-cell nuclear features, it does not assess contextual features that are useful in making a visual diagnosis. Of particular importance are the contextual features of clump thickness, the adherence of cells at the periphery of clumps, and the presence of intact single cells rather than only bare nuclei (14).

The accuracy of the system probably can be increased by combining the new 192 samples with the original 569 samples and retraining. However, any system that is to be used at multiple sites should be trained on samples obtained and analyzed by multiple investigators located at different institutions. Presently, Xcyt has been validated for one investigator located at a single institution. Xcyt currently runs under Linux on a PC, thus obviating the need for a work station.

We conclude that digital image analysis coupled with machine-learning techniques can accurately diagnose breast FNAs, and that the most effective approach combines automated diagnosis with visual diagnosis.

Acknowledgments

This work was supported in part by National Institutes of Health INRSA Fellowship 1 F32 CA 68690-01 (W. Nick Street), Air Force Office of Scientific Research grant AFOSR F49620 (Olvi L. Mangasarian), and National Science Foundation grant CCR-9322479 (Olvi L. Mangasarian).

REFERENCES

1. Stone M. Cross-validatory choice and assessment of statistical predictions. *J Royal Statistical Society* 1974;36:111-47.
2. Kass M, Witkin A, Terzopoulos D. Snakes: active contour models. *Int J Comput Vision* 1988;1:321-31.
3. Street WN, Wolberg WH, Mangasarian OL. Nuclear feature extraction for breast tumor diagnosis. *Proc IS&T/SPIE Int Sympos Electronic Imag* 1993;1905:861-70.
4. Wolberg WH, Street WN, Mangasarian OL. Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates. *Cancer Lett* 1994;77:163-71.
5. Mangasarian OL, Street WN, Wolberg WH. Breast cancer diagnosis and prognosis via linear programming. *Operations Res* 1995;43:570-77.
6. Bennett KP. Decision Tree Construction via Linear Programming. In: Evans M, ed. *Proceedings of the Fourth Midwest Artificial Intelligence and Cognitive Science Society Conference*, Utica, IL: Homestead, AL: Intergraph, 1992:97-101.
7. Mangasarian OL. Mathematical programming in neural networks. *ORSA J Comput* 1993;5:349-60.
8. Breiman L, Friedman J, Olshen R, Stone C. *Classification and Regression Trees*. Pacific Grove, CA: Wadsworth, 1984.
9. Quinlan JR. *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann, 1993.
10. Palombini L, Fulciniti F, Vetrani A, et al. Fine-needle aspiration biopsies of breast masses. A critical analysis of 1956 cases in 8 years (1976-1984). *Cancer* 1988;61:2273-77.
11. Vetrani A, Fulciniti F, de Benedetto G, et al. Fine-needle aspiration biopsies of breast masses. An additional experience with 1153 cases (1985 to 1988) and a meta-analysis. *Cancer* 1992;69:736-40.
12. Painter RW, Clark WEI. Negative findings on fine-needle aspiration biopsy of solid breast masses: patient management. *Am J Surg* 1988;155:387-90.
13. Giard RWM, Hermans J. Fine needle aspiration cytology of the breast with immediate reporting of the results. *Acta Cytol* 1993;37:358-60.
14. Wolberg WH, Tanner MA, Loh W, Vanichsetakul N. Statistical approach to fine needle aspiration diagnosis of breast masses. *Acta Cytol* 1987;31:747-41.