Ultrasound in Med. & Biol. Vol. 14, Sup. 1, pp. 81-87, 1988
Printed in the U.S.A.
0301-5629/88 $3.00 + .00

QUANTITATIVE EVALUATION OF REAL-TIME ULTRASOUND FEATURES OF THE BREAST

JEFFREY C. BAMBER, LEANORE DE GONZALEZ, DAVID O. COSGROVE, and PHILLIP SIMMONS.

Departments of Physics and Nuclear Medicine, Institute of Cancer Research:Royal Marsden Hospital, Downs Road, Sutton, Surrey, SM2 5PT, England.

and

J. DAVEY and J.A. McKINNA

The Early Diagnostic Centre, Royal Marsden Hospital, Fulham Road, London, England.

Abstract - Semi-quantitative diagnostic features were extracted by a visual analysis of the echographic images of selected cases of breast disease and the results stored in a computer database. The long term aim is to create an environment suitable for the use of multivariate statistical methods systematically to evaluate ultrasound interpretive criteria and diagnostic performance in relation to factors such as scanning instrumentation and other diagnostic techniques. Eventually it is hoped that it will be possible to generate a system for computer assisted diagnosis and training. The results of this pilot study serve to demonstrate the feasibility of the approach and a univariate analysis is used to provide a preliminary ranking of diagnostic features. Features found to be particularly valuable for distinguishing benign from malignant solid lesions were the regularity and definition of the edge of the tumour, the mobility of the tumour and measures of echo heterogeneity within and posterior to the tumour mass.

Key Words: Breast, B-scan, Computer Database, Diagnostic Features, Differential Diagnosis, Discriminant Analysis, Echography, Expert System, Multivariate Analysis, Neoplasm, Real-time, Ultrasound, Univariate Analysis.

INTRODUCTION

The value of ultrasound in the diagnosis of breast cancer is now widely recognised, although its reported role has varied from being regarded as a method useful for little more than differentiating between cystic and solid masses (e.g. Cole-Beuglet et al, 1975; Laing, 1976) to a diagnostic technique which can rival mammography in accuracy for distinguishing benign from malignant lesions (e.g. Kobayashi et al, 1985; Pluygers et al, 1977; Ueno et al, 1986). Correspondingly, considerable variability exists in the figures reported for sensitivity and specificity of the method, which might have arisen from many sources, including different population biases, different image observer decision thresholds (observer bias), varying degrees of expertise, different ultrasonic equipments or different examination techniques. There is virtually no information which may be used to discover why different workers have widely different degrees of success in using the method.

The diagnosis of breast disease from ultrasonic images, like mammography, involves visual feature extraction and pattern recognition. The confusion associated with the variability in the success of using ultrasound for breast cancer diagnosis is compounded by a great variation in the diagnostic features which different workers describe as being useful. It is apparent from the literature that some workers are ignoring diagnostic features which others find useful but, at a more subtle level, there is rarely the quantitative information from which to gain a good appreciation of the relative importance of the different image features. Variation in the scanning instrumentation and technique used might account for some of these differences but is not clear to what degree specific diagnostic features are better appreciated using one type of instrument or examination technique compared to another.

It would appear that a more suitable framework from which to perform systematic studies of diagnostic features is called for, to provide:

(a) a basis for evaluating quantitatively the relative importance of each of the diagnostic features alone and in combination,

(b) an objective means of comparing results from different instruments,

(c) a means for evaluating the degree of improvement resulting from new developments in the instrumentation such as enhanced resolution (Foster et al, 1985) or speckle reduction processing (Bamber, 1985), and

(d) a basis for evaluating the contribution of, and for understanding the nature of any additional information obtained by the addition of new methods of examination such as ultrasound Doppler observation of blood flow in the breast (Bamber et al, 1983; Burns et al, 1982; Jellins, 1985).

It was decided that these aims might be achieved by using a computer-based compository of data on visually extracted diagnositic features, patient information and histological findings, in combination with the use of multivariate statistical methods to provide the pattern recognition process necessary for diagnosis. Such a system would permit, for example, a full assessment of diagnostic accuracy in which an ROC (receiver-operating characteristic) curve can be generated by varying the computer decision threshold. ROC curves provide an adequate objective method for performing the evaluations mentioned in (a) to (d) above. Single estimates of sensitivity or specificity do not.

In the long term, computer classification of cases into disease classes based on statistical pattern recognition could be used to aid diagnosis (computer assisted diagnosis, or CAD) and to help with training in the use of ultrasound in diagnosing breast disease (computer assisted training, or CAT). This paper, however, reports a limited, short-term pilot study. Ultrasound examination of the breast had only been introduced to our breast cancer early diagnostic centre shortly before this study was undertaken. As relatively inexperienced users of the method we felt that the process of deciding which features to record, and the discipline of having to perform an objective visual analysis of the images before entering the features into a computer, would greatly help us learn to make effective use of this new tool.

MATERIALS AND METHODS

Ultrasound Equipment

Initially, and for all of the cases reported in this study, a Hitachi EUB-26 real-time scanner was used, equipped with a wide field of view (11.5 cm) 5 MHz linear array transducer. Later, this has been changed to an Aloka SSD-121, mechanical sector system operating at 7.5 MHz. It has been found worthwhile to trade the wide field of view of the linear array for the narrower (40 degree) sector but superior resolution of the higher frequency probe. Both scanners are straightforward, non-specialist but relatively cheap systems, already found in many hospitals.

Patients and examination technique

For use in this pilot study patients (95 in all) were selected for whom the decision had been made (on the grounds of symptoms, clinical examination or mammogram) that they required excision biopsy or aspiration cytology. The final diagnosis, for correlation with the ultrasound findings, was taken from histology or, if this was not available, from cytology. This patient population included 28 carcinomas, 53 benign solid masses, 11 cysts and 3 normal subjects, although the analysis applies only to evaluating the ability of ultrasound features to discriminate between benign and malignant solid masses.

The examination technique was conventional for contact scanning; both breasts were scanned sequentially with the patient in the supine position, slightly rotated away from the side being examined and with the arm elevated.

Feature extraction

The features used were arrived at by attempting to compile and, to some degree, extend those published as having been used by other workers. To provide a basis for future equipment comparison, features used by diagnosticians working with large water-path scanners were retained even though some may not have been thought to be applicable to images obtained with contact real-time scanners. Diagnostic features were classified into those relating to the general structure of the breast and those directly relating to the masses, if present. Some of the general breast features have previously been termed secondary features diagnostic of breast cancer (Jellins et al, 1985). They relate a discrete lesion to the breast as a whole, in terms of anatomical distortion and disturbance. Examples are skin thickening, local loss of subcutaneous fat space, and distortion of suspensory (Cooper's) ligaments. Others, such as the proportion of the breast occupied by subcutaneous fat and the overall glandular pattern, are not the secondary features referred to by others but are of potential interest because of possible correlation with breast cancer incidence. The "mass features" have been termed primary diagnostic features by others. These are broadly of four types, although strict subclassification at this level has not been made in the database:

(a) morphological (e.g. size, shape, regularity and smoothness of borders, presence of an echogenic zone surrounding the mass, heterogeneity of the echo pattern within specific regions),

(b) acoustic properties (echogenicity within the mass, attenuation due to the mass),

(c) deformability and mobility of the mass assessed by ultrasound observation under deliberate mechanical distortion of the tissues (ultrasound palpation),

(d) criteria derived by making comparisons between features, possibly between ultrasound features only but also between ultrasound features and those obtained by other means (e.g. ratio of the size of a mass as apparent from clinical palpation to that apparent as the echo-poor region on the ultrasound image).

The degree of presence or absence of each feature was assessed by the examiner (LdeG) using both categorical and continuous scoring systems, as appropriate to the feature in question. An important point is that feature assessment was always based on a real-time search through the volume of the tissue and never on the appearances within a single image. Thus, for example, heterogeneity of the echo pattern or shadowing refers to the heterogeneity from image to image, as well as within each image.

The database

For the series of cases reported in this paper, the database and computer program written for data entry were constructed using the dbaseII* language in the CPM** operating system. Features and other details were entered interactively by filling in forms generated on the computer screen. This system was later found to have severe limitations, particularly with regard to making changes to the database schema (such as adding new features) and when interrogating the database or reducing the files to a form suitable for statistical analysis. A new version of this database is now under construction which overcomes some of the limitations mentioned above. The main features of the new system are: (a) it is based on an IBM Personal Computer for low cost and wide availablity, (b) a relational data model is used to give the system flexibility to evolve and to simplify maintainance, (c) the system chosen makes use of standard SQL (structured querie language) for programming and low level database interrogation. This system will be the starting point for the main phase of this project, which we will report on at a later date.

* Trade name of Ashton Tate. ** Trade name of Digital Research.

Analysis

To assess the relative value of the visually scored features and to explore methods for combining them to distinguish benign from malignant masses, the data were transferred to a VAX minicomputer where the SAS* package was used to perform the calculations.

RESULTS

With the small number of patients involved at this stage of the study a thoroughly meaningful and complete analysis of the data is not possible, particularly using multivariate methods. We therefore report preliminary results only and concentrate on the mass features referred to earlier. For all of the results reported below the cysts were excluded, i.e. the analysis relates only to the benign and malignant solid masses. The results serve to demonstrate that the graded scoring systems used for the features are in most cases acceptable and useful, and provide some indication of the relative value of the different features for diagnostic purposes, when each is used alone.

General points

In this series cancers tended to be greater than 1.2 cm diameter, whereas a substantial proportion of the benign masses were smaller than this. No account has been taken of any possible variation of the incidence of diagnostic features with the size of the lesion but it is anticipated that, as the number of cases increases, it will be of interest to investigate this in the future. About 41% of all masses occured in the upper outer quadrant of the breast, there being no significant difference between benign and malignant masses in this respect.

Univariate analysis of mass features

Figure 1 shows the distributions for 10 mass features for which categorical scoring systems had been devised, from classes (a), (b) and (c) mentioned in the section on feature extraction. Skin attachment was also in the list of mass features but was absent for all of the masses in this study. Mass features with continuous scoring systems amounted largely to the measurement of the three echographic diameters (defined as the echo poor central region, or nidus), both before and after compression (applied with the transducer), and two clinical measurements of diameter (transverse and longitudinal) obtained by palpation. These then led to the derived features of catagory (d), as follows:

Shape#1 = The transverse echographic diameter divided by the longitudinal echographic diameter

Shape#2 = The maximum of either the longitudinal or transverse echographic diameters divided by the depth diameter

Shape difference = The percentage difference in the value of shape#2 before and after compression.

Size difference = The percentage difference between the average clinical (palpable) diameter and the average of the longitudinal and transverse echographic diameters.

For each feature a two-way table was constructed between the feature score and the diagnostic class (benign or malignant). The Pearson chi-squared non-parametric measure of association between two variables (SAS, 1985) was then used for assessing the value of each feature, against the null hypothesis that there is no significant difference between the feature values for malignant and benign masses. Unfortunately, because it was time consuming to make the continuous measurements, the number of patients for which the derived features mentioned above were obtained was too small for a valid chi-squared test to be performed. Therefore, although no significant differences were observed between the values of these features for benign and malignant masses, it remains of interest to collect a larger data set. The shape factors are similar to the so-called L/T ratio diagnostic features found to be useful by some Japanese workers (e.g. Tajima et al, 1983). The shape difference is an attempt to quantify compressibility whereas the size difference between palpable and echographic measurements may be large for

* Trade name of SAS Institute.

malignant masses (Bamber and Yarnold, 1983) and has been found to be diagnostically useful (Dale and Gros, 1980; Nishimura et al, 1986).

Chi-squared tests were performed for all other features, in some cases grouping the scoring categories so as to compensate for the small number of cases. In total 8 features showed a significant difference between benign and malignant populations (defined as the probability that the null hypothesis is correct is less than 0.01), including one of the general breast features. These are shown in table 1, along with those which showed poor diagnostic value.

Multivatiate analysis

The chi-squared analysis used above provides a measure of the relative diagnostic usefulness of each feature when used in isolation. Diagnosic judgements, however, rely strongly on combinations of features; i.e. the assessment of the probability of a number of features all possessing particular values simultaneously. It is quite possible, for example, that two features will provide an accurate diagnosis when used in combination even though they are of no value when they are used alone. A strong component of the motivation for this study was to apply methods of multivariate statistics to the data and, for this reason alone, it is important to continue to study all of the features, even those which appeared to possess little value when assessed by univariate methods.

The data set at present is too small to permit a meaningful multivariate analysis. A preliminary application of a stepwise discriminant analysis did however produce a ranking of the most valuable features which included many of the features listed on the left of table 1 and also included the derived feature, shape#2. The discriminant analysis program available uses parametric methods and assumes normally distributed continuous feature values. Application of such discriminant procedures to binary and categorical features, as used here, is common (Gale et al, 1984) but can lead to misleading results (Goldstein and Dillon, 1978).

Table 1. Grouping of diagnostic features according to the probability that the null hypothesis (i.e. there being no difference between benign and malignant populations) is true. Some features are missing from this table because the small sample size prevented a valid statistical test.

Features with probability < 0.01 (high diagnostic value)	Features with probability > 0.01 (relatively poor diagnostic value)
Edge regularity	Shape of mass
Edge definition	Compressibility
Mobility	Internal echo level
Posterior echo heterogeneity	
Internal echo heterogeneity	
Posterior echo level (shadowing)	
Border echogenicity	
Breast type (general pattern)	

CONCLUSION

For this pilot study a set of diagnostic features was compiled by reviewing the literature. A preliminary scoring system for each feature was defined, resulting in a mixture of binary, categorical and continuous variables. Such a scoring system was shown to be feasible and experience was gained of entering the values into a computerised database. Most of the published mass features were demonstrated and an initial univariate analysis of the results was used to order the features according to their relative usefulness. Features which appeared to be particularly valuable were the regularity and definition of the edge of the tumour, the mobility of the lesion and measures of echo heterogeneity both within and posterior to the tumour mass. We recommend this type of study, where diagnostic criteria and incidence figures are quantitatively evaluated, to any group (like ourselves) just starting to use ultrasound in breast examinations.

A combination of features is required for diagnosis. When more data are available it should be appropriate to apply multivariate statistical methods to provide computer aided systems of diagnosis and learning. It would seem advantageous to speed up the process of data collection by international collaboration. The database is currently being revised to make such collaboration and pooling of data easier to implement, but it will also be necessary and important to standardize on definitions of features and on methods for extracting them.

ACKNOWLEDGEMENTS

This work is supported by grants from the Medical Research Council and the Cancer Research Campaign.

REFERENCES

Cole-Beuglet, C., Soriano, R., Pasto, M., Baltorowich, O., Rifkin, M., Kurtz, A. and Goldberg, B. (1983) Solid breast mass lesions: can ultrasound dirrerentiate benign and malignant? In *Ultrasonic Examination of the Breast* (Edited by J. Jellins and T. Kobayashi) pp. 45-55. John Wiley and Sons, Chichester.

Bamber, J.C. (1985) Improving image quality in breast echography by speckle reduction. In *Proc. Fourth Int. Congress on the Ultrasonic Exam. of the Breast* (Edited by J. Jellins, G. Kossoff and J. Croll) pp. 205-213. International Congress on the Ultrasonic Examination of the Breast, Department of Surgery, Royal North Shore Hospital, St. Leonards, Australia.

Bamber, J.C. and Yarnold, J.R. (1983) Breast tumor measurements using ultrasound. *op. cit.* Cole-Beuglet et al (1983) p.359.

Bamber, J.C., Sambrook, M., Minasian, H. and Hill, C.R. (1983) Doppler study of blood flow in breast cancer. *op. cit.* Cole-Beuglet et al (1983) pp. 371-378.

Burns, P.N., Halliwell, M., Wells, P.N.T. and Webb, A.J. (1982) Ultrasonic Doppler studies of the breast. *Ultrasound Med. Biol.* 8 127-143.

Dale, G. and Gros, CH. (1980) Tissue characterization in breast echography. In *Ultrasonic Tissue Characterization* (Edited by J.M. Thijssen) pp.16-25. Stafleu's Scient. Publ. Co., Alphen aan den Rijn/Brussels.

Foster, F.S., Fishell, E.K., Khodai, M., Harasiewicz, K. and Hunt, J.W. (1985) Initial clinical performance of a prototype breast imaging system using a conical/annular array hybrid transducer. *op. cit.* Bamber (1985), pp. 185-192.

Gale, A.G., Howe, E., Roebuck, E.J. and Worthington, B.S. (1984) Computer aided diagnosis in mammography. In *Computers in Radiotherapy and Oncology* pp.123-128. Adam Hilger, Bristol.

Goldstein, W.R. and Dillon, M. (1978) *Discrete Discriminant Analysis* John Wiley and Sons, New York.

Jellins, J. (1985) B-mode and Doppler assessment of breast disease. *op. cit.* Bamber (1985), pp. 215-220.

Jellins, J., Boyd, J., Kossoff, G. and Reeve, T.S. (1985) Primary and secondary features of solid breast lesions. In *Proc. 4th Meeting of the World Fed. for Ultrasound in Med. Biol.* (Edited by R.W. Gill and M.J. Dadd) pp.303-331. Pergamon, Sydney.

Kobayashi, T., Hayashi, M. and Arai, M. (1985) Echographic characteristics and ultrasonic tissue characterization in breast tumor (review). *J. UOEH* 7 419-434.

Laing, F.C. (1976) Ultrasonographic evaluation of breast masses. *J. de l'Assoc. Canadienne des Radiologistes* 45 278-282.

Nishimura, S., Kashihara, S. and Matsusue, S. (1986) Clinical significance of the ratio of breast tumor diameter by palpation to that by ultrasonography. In *J. Soc. Ultrasound Med. Proceedings*, May. 1986, 1423-1427.

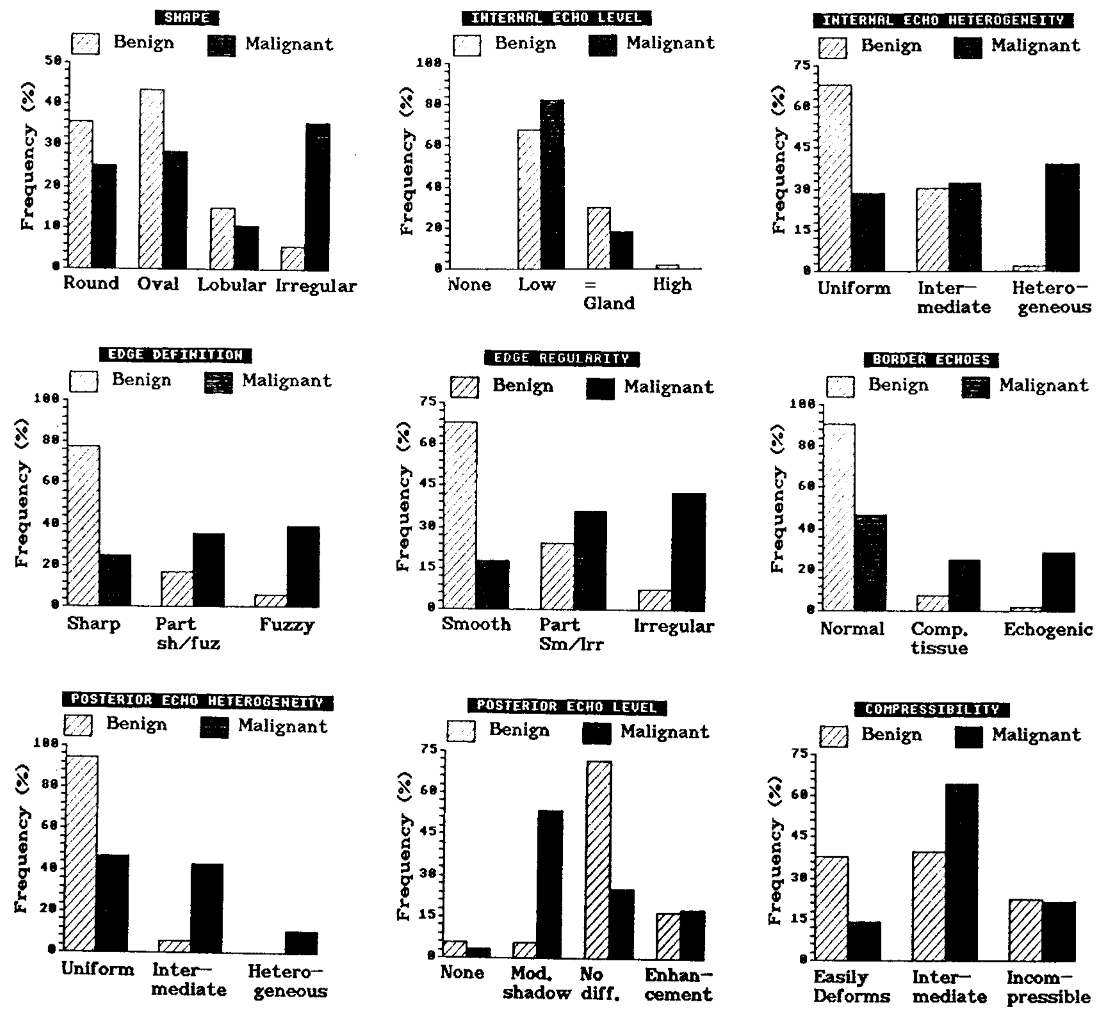
1426.

Pluygers, E., Rombaut, M., Beauduin, M., Hendrickx, B., Devergnies, F., Pieron, A. and Tamignaux, G. (1977) Possibilities, results and indications of breast echography. *J. Belge Radiol.* 60 181-196.

SAS (1985) *SAS User's Guide: Statistics* Version 5 Edition, p.412. SAS Institute Inc., Cary, N.C.

Tajima, T., Kubota, M., Mitomi, T., Nanri, K., Sakurai, I. and Kobayashi, H. (1983) Longitudinal/transverse ratio of tumor echogram as a diagnostic criterion of breast carcinoma. *op. cit.* Cole-Beuglet et al (1983) pp. 69-70.

Ueno, E., Tohno, E. and Itoh, K. (1986) Classification and diagnostic criteria in breast echography. *Jpn. J. Med. Ultrasonics* 13 19-31.



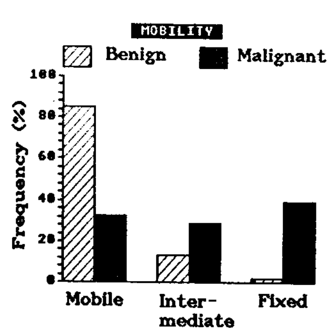


Figure 1. The frequency distributions of ten "mass features" for benign and malignant solid breast lesions. Frequency is the number of occurrences of a particular feature value divided by the total number of cases within a diagnostic class, whether benign or malignant.