INT. J. SYSTEMS SCI., 1991, VOL. 22, NO. 10, 1985–1991

Methods for classification of two thyroid follicular tumour classes

S. K. PARUI†‡, U. JUETTING‡ and G. BURGER‡

The goal of the present paper is to show that certain cytometric (morphological, photometric and textural) features of isolated cell nuclei can be useful for discrimination and classification of two thyroid tumour classes, namely, follicular adenoma and follicular carcinoma which cannot in general be visually discriminated in cytological smears. Several linear classifiers both at cell and specimen levels are proposed. In order to estimate the true error rate of these classifiers, a 'hold one specimen out' scheme is employed.

1. Introduction

Important goals in clinical cytology are cancer diagnosis, cancer typing, prognostic grading, etc. These tasks are patient oriented and final decision making refers to specimens and not to cells. One approach to these tasks is through certain quantitative features of isolated cell nuclei (Burger *et al.* 1985, Burger and Juetting 1986). In this case, first the cell images in a specimen are segmented and then a set of morphological, photometric and textural features are extracted from the nucleus of each cell. Specimen classification may then be a two-stage scheme. In the first stage, each of the cells in a specimen is classified into one of the member classes. In the second stage, on the basis of the first stage classification results, the specimen itself is classified into one of the specimen classes. The two-stage specimen classification problem has been considered first by Castleman and White (1980, 1981) and later by several others (Burger and Juetting 1986, Timmers 1987, Smeulders 1986). The present paper deals with two cytopathologically relevant thyroid tumour classes where a fine needle aspirate specimen is to be classified into one of them. It is assumed that there are two classes at cell level also. We employ statistical linear discriminant analysis to obtain an optimal classifier at the cell level and on the basis of that devise a specimen classifier which is not necessarily optimal at the specimen level.

Also proposed is a specimen classification approach that does not involve cell classification but is based only on the cell features mentioned above. The two thyroid classes that are dealt with here are follicular adenoma and follicular carcinoma which in general are not visually distinguishable in cytological smears. It is assumed that an overwhelming majority of cells in the carcinoma specimens are carcinoma cells, though they cannot be identified. Now, since a pure carcinoma cell class cannot be formed, the pooled cells from all carcinoma specimens are treated as the carcinoma cell class. This is later proved to be not so unreasonable by the results of a 'hold one specimen out' scheme where a cell classifier is devised using all but one specimen and then the cells in that specimen are classified by the classifier. The cells that are classified as 'carcinoma' are called positive. Specimen classification is done on the basis of the positive cells. The adenoma cell class is formed by pooling all the cells from all adenoma specimens.

Received 27 March 1990.

† Guest Scientist from the Electronics and Communication Sciences Unit, Indian Statistical Institute, 203 B.T. Road, Calcutta, 700035, India.

‡ Institut für Strahlenschutz, G. S. F. München, D-8042 Neuherberg, Germany.

It is well known that if the training set is used as a testing set for a classifier, the estimate of the misclassification rate is optimistically biased (Hand 1981). In order to solve this problem the 'hold one specimen out' scheme is employed to achieve a more reliable estimate of the misclassification rate. It is finally shown through numerical results how reduction in dimensionality may lead to reduced true misclassification rates.

2. Cytological material and feature set

The present analysis is based on 27 follicular adenoma and 20 follicular carcinoma specimens which were obtained by fine needle aspiration before surgery and on the diagnosis that was done by a pathologist after surgery. The smears were air dried and pappenheim stained. From each specimen, about 100 isolated well preserved cell nuclei were visually selected for analysis without any pathological biasing. The images of the cells were captured by a TV camera with microscopical objective magnification of 100 X (oil) using a narrow band optical filter at 500 nm wavelength, and digitized. The nominal local resolution of the digitized images was 0·25 μm and the nominal gray value range 256. Using ILIAD procedures (Eriksson *et al.* 1982) which are interfaced with a VAX 11/750 (Gais and Rodenacker 1987), 43 features are extracted from each cell image. These features include (*a*) morphological features such as area and perimeter, (*b*) photometric features such as integrated optical density and (*c*) textural features describing the staining pattern of a cell. Details of the features are available elsewhere (Rodenacker 1987).

At the cell level the set of cells pooled from all 27 adenoma specimens form the sample adenoma cells. The sample set of carcinoma cells is formed similarly. These two sample sets were used to devise the cell level classifier. The classification here is supervised and uses linear discriminant analysis.

3. Feature reduction and cell classification

To use all the 43 features in discrimination and classification is not practical for two reasons. First, to compute all the 43 features for each cell in an unknown specimen is very expensive in terms of computing time. Second, the accuracy of a classifier may, beyond a level, decrease with increasing number of features; this is known as Bellman's 'curse of dimensionality' (Hand 1981). This curse of dimensionality is demonstrated in a later section.

It is seen that there is a high amount of redundancy in the set of 43 features with respect to discriminatory power. As a first step to reducing the number of features in the classifier, a smaller number of features from the original set of 43 features were selected. For this purpose, a stepwise discriminant analysis was employed.

The stepwise method selects only the features with significant extra discriminatory power and the number of computational steps involved in the process depends on the data set. The method starts with no features in the model. At each step, if the model is non-empty, the feature in it that contributes least to the discriminatory power of the model as measured by Wilks's lambda is considered. If this contribution is not significant (in terms of a preassigned value) the feature is removed from the model. Otherwise, irrespective of whether the model is empty, the feature not in it that contributes most to the discriminatory power of the model is considered. If its contribution is significant (again in terms of a pre-assigned value), then the feature is included into the model. When each feature in the model contributes significantly to

	Adenoma (%)	Carcinoma (%)	Total (%)
(*a*) Using all 43 features. The misclassification rate is 33·21% (1494)			
Adenoma	1673	843	2516
	(66·49)	(33·51)	(100)
Carcinoma	651	1331	1982
	(32·85)	(67·15)	(100)
(*b*) Using only 13 best features. The misclassification rate is 34·17% (1537)			
Adenoma	1629	887	2516
	(64·75)	(35·25)	(100)
Carcinoma	650	1332	1982
	(32·80)	(67·20)	(100)
(*c*) Results using only the canonical variable are the same as in (*b*)			

Table 1. Cell level reclassification results using linear discriminant analysis with equal *a priori* probabilities. There are $N_1 = 2516$ cells in 27 adenoma specimens and $N_2 = 1982$ cells in 20 carcinoma specimens.

the discriminatory power of the model and no feature outside the model does so, the stepwise selection procedure stops.

In the present case, both the criteria above are made stringent so that not too many features are finally selected in the model. In terms of the F test the significance level is taken to be as low as 0·005 in both cases. The final model obtained through this stepwise selection consisted of only 13 features. The loss in the classification results at the cell level due to the removal of 30 features is insignificant and shown in Table 1.

Thus in the first phase of feature reduction, a subset of 13 features from the original feature set is selected. But 13 is considered still too high for the dimensionality of the cell classifier. In the second phase of feature reduction, the aim is not to reduce the cost of measurement of features but to increase the accuracy of the classifier (Hand 1981). For this purpose the canonical discriminant analysis is used which derives a linear combination of the 13 features (the first canonical component); this gives the highest between class variation. This linear combination is called the canonical variable. This variable alone is used in the cell classifier. The increase in accuracy achieved by replacing 13 features with the canonical variable in the cell classifier can be seen later (in Tables 3 and 4).

4. Specimen classification

4.1. *Reclassification scheme*

So far the classification has been confined to the cell level. The problem now is to classify a specimen on the basis of its classified cells. Here it is assumed that each specimen class is characterized by the proportion p of positive cells present in a specimen and that the number of these cells follows a binomial distribution Bin (n, p) where n is the total number of cells in the specimen. Let p_1 be the population parameter for this proportion in the adenoma specimens and p_2 for this proportion in the carcinoma specimens. Let p_1^* be the ratio of the total number of positive cells present in 27 adenoma specimens to the total number of cells in them. Similarly, p_2^* is computed on the basis of 20 carcinoma specimens. p_1^* and p_2^* are the maximum likelihood estimators of p_1 and p_2, respectively. It can be shown that p_1^* follows an asymptotic normal distribution with mean p_1 and variance $p_1^*(1 - p_1^*)/N_1$. Similarly,

	Adenoma	Carcinoma	Unclear	Total
(*a*) Using all the 43 features. Number of correctly classified specimens is 33 (78·57%). Number of incorrectly classified specimens is 9 (21·43%)				
Adenoma	18	6	3	27
Carcinoma	3	15	2	20
(*b*) Using the best 13 features. Number of correctly classified specimens is 32 (78·05%). Number of incorrectly classified specimens is 9 (21·95%)				
Adenoma	17	6	4	27
Carcinoma	3	15	2	20
(*c*) Results with only the canonical variable are the same as in (*b*).				

Table 2. Specimen level classification results with reclassification scheme corresponding to the cell level classification results shown in Table 1. The percentages exclude unclear specimens.

p_2^* follows an asymptotic normal distribution with mean p_2 and variance $p_2^*(1 - p_2^*)/N_2$. (N_1 and N_2 are the total numbers of cells present in the adenoma and carcinoma specimens, respectively). Let these two distributions be denoted by f_1 and f_2, respectively (p_1 and p_2 being replaced by samples p_1^* and p_2^*, respectively). Let t be the threshold value for the positive cell proportion such that $f_1(t) = f_2(t)$ and t lies between p_1 and p_2. Now the specimen classification problem reduces to deciding which of the above two populations a specimen with n cells (among which x are positive cells) is coming from. Let p be the population ratio of positive cells in such a specimen. Then p^* follows an asymptotic normal distribution with mean p and variance $p^*(1 - p^*)/n$ where $p^* = x/n$.

Note that the 90% confidence interval for p is $(p^* - 1{\cdot}645\sqrt{(p^*(1 - p^*)/n)}, p^* + 1{\cdot}645\sqrt{(p^*(1 - p^*)/n)})$. If $p^* + 1{\cdot}645\sqrt{(p^*(1 - p^*)/n)}$ is less than t, then the specimen is classified as adenoma. On the other hand, if $p^* - 1{\cdot}645\sqrt{(p^*(1 - p^*)/n)}$ is greater than t, then the specimen is classified as carcinoma. Otherwise, the specimen is not classified but left as unclear. That is, a specimen is not classified if t falls inside the above confidence interval. The specimen level classification results using this scheme are shown in Table 2.

4.2. *Hold one specimen out scheme*

In Tables 1 and 2, the results are of reclassification in the sense that the same set is used for both learning and testing. To obtain the true error (or misclassification) rate one should use a different testing set from the learning set. But in the present case (and in the medical field in general), not enough specimens are normally available which can be divided into a learning set and a testing set. For this reason, a hold one specimen out scheme is employed. That is, when a specimen is being classified, it is excluded from the learning set while all other specimens are included, and the excluded specimen becomes the testing set. This is repeated for each specimen and the average proportion of misclassified cases is an estimate for the true misclassification rate. But is should be noted that two things are learnt on the basis of all the specimens together: first, the set of 13 most discriminatory features and second, the coefficients of the canonical variable. These are learnt only once and learnt before employing the hold one specimen out scheme. The results of three different classifiers at both cell level and specimen level using this scheme are given in Tables 3 and 4, respectively.

In specimen level classification also, the fact that less number of features can be more accurate is evident. For the classifier with all 43 features, results with the hold

	Adenoma (%)	Carcinoma (%)	Total (%)
(*a*) Using all 43 features. The misclassification rate is 45·40% (2042)			
Adenoma	1398	1118	2516
	(55·56)	(44·44)	(100)
Carcinoma	924	1058	1982
	(46·62)	(53·38)	(100)
(*b*) Using 13 best features. The misclassification rate is 42·49% (1911)			
Adenoma	1447	1069	2516
	(57·51)	(42·49)	(100)
Carcinoma	842	1140	1982
	(42·48)	(57·52)	(100)
(*c*) Using only the canonical variable. The misclassification rate is 34·24% (1540)			
Adenoma	1627	889	2516
	(64·67)	(35·33)	(100)
Carcinoma	651	1331	1982
	(32·85)	(67·15)	(100)

Table 3. Cell level classification results with hold one specimen out scheme using linear discriminant analysis. All three tables below show true error rates. The results demonstrate how a classifier with more features can sometimes be less accurate (Bellman's curse of dimensionality).

one out scheme are far worse than those with the reclassification scheme. For the classifier with 13 features, the difference is less. Finally with the canonical variable the difference is nil. Thus, the instability of a classifier increases with larger number of features and this can be explained in the following way. For a classifer with more features, more number of parameters are to be estimated. But the number of observations on the basis of which this estimation is to be made remains the same. In other words, the same set of observations if seen in a higher dimensional space is quite likely to be sparsely distributed and hence will lead to unstable and unreliable estimates of the parameters. Consequently, the performance of a classifier with more features becomes worse.

The specimen classifier above classifies each cell of a specimen. An alternative way is not to classify the cells which lie very close to the boundary between the two classes.

	Adenoma	Carcinoma	Unclear	Total
(*a*) Using all the 43 features. Number of correctly classified specimens is 24 (61·54%). Number of incorrectly classified specimens is 15 (38·46%)				
Adenoma	12	11	4	27
Carcinoma	4	12	4	20
(*b*) Using best 13 features. Number of correctly classified specimens is 25 (65·79%). Number of incorrectly classified specimens is 13 (34·21%)				
Adenoma	13	9	5	27
Carcinoma	4	12	4	20
(*c*) Using only the canonical variable. Number of correctly classified specimens is 32 (78·05%). Number of incorrectly classified specimens is 9 (21·95%)				
Adenoma	17	6	4	27
Carcinoma	3	15	2	20

Table 4. Specimen level classification results with hold one specimen out scheme corresponding to the cell level classification results shown in Table 2. The percentages exclude unclear specimens.

	Adenoma (%)	Carcinoma (%)	Unclear (%)	Total (%)
(*a*) Cell level classification results				
Adenoma	1431	591	494	2516
	(56·88)	(23·49)	(19·63)	(100)
Carcinoma	483	1116	383	1982
	(24·37)	(56·31)	(19·32)	(100)
(*b*) Specimen level classification results				
Adenoma	20	6	1	27
Carcinoma	3	16	1	20

Table 5. Cell and specimen level classification results with hold one specimen out scheme on the basis of new cell classification approach where only a cell with *a posteriori* probability more than 0·55 is classified. Only the canonical variable is used.

This may be appropriate since the overlap between the two cell classes is quite high. A cell is classified if the maximum of the two *a posteriori* probabilities is greater than 0·55. Otherwise, the cell is left unclassified. Then the same specimen classification technique is applied as is described for the first specimen classifier above based on the classified cells of a specimen. The classification results in the cell and specimen levels are given in Table 5. It can be seen that four out of six specimens unclassified before are now classified correctly.

The specimen classification discussed so far involves, directly or indirectly, single cell classification. Now a specimen classification scheme is proposed which avoids single cell classification altogether and is based on the mean value of the canonical variable V within a specimen. It is assumed that V is normally distributed within adenoma specimen cells and within carcinoma specimen cells with different means but equal variance. Thus, the average of the two means is taken as the threshold value T. It is seen that the adenoma mean is greater than the carcinoma mean. Now, suppose a specimen with n cells and mean $\bar{v}$ is to be classified. The standard error is $s'' = s/\sqrt{n}$. Thus, the sample or specimen mean $\bar{V}$ is normally distributed with mean $\bar{v}$ and standard deviation s''. If $\text{Prob}(\bar{V} > T)$ is greater than 0·95, that is, if $\bar{v} - 1{\cdot}645s''$ is greater than T, then the specimen is classified as adenoma. On the other hand, if $\text{Prob}(\bar{V} < T)$ is greater than 0·95, that is, if $\bar{v} + 1{\cdot}645s''$ is less than T, then the specimen is classified as carcinoma. Otherwise, the specimen is not classified and is left as unclear. The specimen level classification results using this approach are given in Table 6. It can be seen that the specimen classification results based on the specimen mean are more or less the same as those based on single cell classification (Table 4(*c*)).

	Adenoma	Carcinoma	Unclear	Total
(*a*) Using only the canonical variable. Number of correctly classified specimens is 33 (80·49%). Number of incorrectly classified specimens is 8 (19·51%).				
Adenoma	18	5	4	27
Carcinoma	3	15	2	20

Table 6. Specimen level classification results on the basis of the specimen mean of the canonical variable V. Reclassification and hold one specimen out schemes produce the same results.

5. Discussion

The intention of this paper has been to explore the possibility of using certain cell features for the classification of some thyroid tumour classes on the basis of smears from fine needle aspiration without having to go for surgery. The results above indicate that the cell features do contribute to the classification. But it is to be noted that though the cell level classifiers are optimal the specimen level classifiers are not necessarily so. Thus, there is a scope for improvement in the specimen classifiers. It has been observed that within a diagnostic class, the between specimen variability is quite high though it has been implicitly assumed that a specimen can be treated as a random sample from a normal population (of the pooled cells from the corresponding diagnostic class). A more realistic model will possibily be that of a compound distribution where one can assume one distribution within a specimen and another between specimens within a diagnostic class (Timmers 1987, Bartels 1988). For example, it can be assumed that the cell feature vector X follows $N(\mu, \sigma_1^2)$ within a specimen and μ follows $N(\alpha, \sigma_2^2)$ within a diagnostic class, where σ_1^2 reflects the variability within a specimen and σ_2^2 the variability within a diagnostic class.

For the present analysis of data the statistical software package SAS (SAS Institute Inc., U.S.A.) has been used on an IBM 4381-2 machine.

As mentioned earlier, there is no information on which are the carcinoma cells in the carcinoma specimens. In case it is in some way possible to identify the carcinoma cells in a carcinoma specimen, the two thyroid tumour classes under consideration may be discriminated more successfully employing the methods proposed in the present paper.

Acknowledgments

The authors would like to thank Dr Schenck for providing the fine needle aspirate specimens and Dipl. Math. Rodenacker and Dipl. Ing. Gais for extracting the cytometric features from these specimens.

References

Bartels, P. H., 1988, *Morphometrie in der Zyto-und Histopathologie*, edited by G. Burger, M. Oberholzer and W. Gossner (Berlin: Springer-Verlag), p. 243.

Burger, G., and Juetting, U., 1986, *Pattern Recognition in Practice II*, edited by E. S. Gelsema and L. N. Kanal (Amsterdam: North-Holland), p. 509.

Burger, G., Rodenacker, K., Juetting, U., Gais, P., and Schenck, U., 1985, *Acta Stereol*, **4**, 243.

Castlemanm, K. R., and White, B. S., 1980, *Anal. Quantitative Cytol.*, **2**, 117; 1981, *Cytometry*, **2**, 155.

Eriksson, O., Bengtsson, E., Jarkrans, T., Nordin, B., and Stenkvist, B., 1982. *Proc. Int. Symp. on Medical Imaging and Image Interpretation* (ISMII-82), Berlin.

Gais, P., and Rodenacker, K., 1987, *Clinical Cytometry and Histometry*, edited by G. Burger, J. S. Ploem and K. Goerttler (London: Academic press), p. 36.

Hand, D. J., 1981, *Discrimination and Classification* (New York: Wiley).

Rodenacker, K., 1987, *Clinical Cytometry and Histometry*, edited by G. Burger, J. S. Ploem and K. Goerttler (London: Academic press), p. 91.

Smeulders, A. W. M., 1986, *Pattern Recognition in Practice II*, edited by E. S. Gelsema and L. N. Kanal (Amsterdam: North-Holland), p. 497.

Timmers, T., 1987, Pattern recognition of cytological specimens. Ph.D. thesis, Vrije Universiteit te Amsterdam.