**Fine Needle Aspirate of Breast Lesions Dataset**

Dr Simon S Cross, Senior Lecturer, Department of Pathology, University of Sheffield Medical School, Beech Hill Road, Sheffield S10 2UL, UK, s.s.cross@sheffield.ac.uk

Dr Robert F Harrison, Reader, Department of Automatic Control & Systems Engineering, University of Sheffield, Sheffield, UK.

**The problem domain**

Breast cancer is the commonest cancer (excluding skin cancer) which affects women in North America, Europe and the Antipodes. (1) The prognosis for breast cancer is primarily dependent on how far the tumour has spread before treatment is instituted and this is a relatively direct function of time. If the diagnosis of breast cancer can be made earlier then the prognosis is improved since more cases will have disease localised to the breast, without spread to lymph nodes in the axilla or more distant sites. A number of countries, including the UK and USA, have instigated screening programmes for breast cancer which use radiographs of the breast (mammograms) as the screening modality. Mammographic abnormalities which raise the suspicion of malignancy include microcalcification and parenchymal deformity but these are not entirely specific and a confirmatory method of diagnosis is required before definitive therapy can be instituted. (2) The most common confirmatory method is fine needle aspiration of the breast lesion (FNAB) and cytological examination. (3) In this method cells from the breast lesions are sucked into a syringe through a fine bore needle (similar to that used for taking blood samples) and are then transferred to a transport solution and sent to a pathology laboratory. In the laboratory the fluid is spun down in a cytocentrifuge to produce a deposit of cells on a glass slide. (4) This slide is stained, usually using the Papanicolaou or Giemsa methods, and is then viewed down the microscope by a trained cytopathologist.

The process by which cytopathologists make their diagnoses is largely unknown but appears to be mainly one of pattern recognition with occasional use of heuristic logic. (5) Cytopathologists are trained by an apprenticeship process in which they view slides down a double-headed microscope with an expert and are told the expert’s
opinion of the diagnosis. The expert may also point out specific features in the specimen and attribute a qualitative value to them in the diagnosis of either benignancy or malignancy. There are also textbooks and journals that codify this information. The full training of a newly-qualified medical doctor to an independently practising cytopathologist takes a minimum of 5 years. There have been many studies of the accuracy of FNAB cytodiagnosis which has been shown to be high in specialist centres but much lower in non-specialist centres when the technique is first introduced. There is thus considerable scope for decision aids which could accelerate the training process or assist in diagnosis in non-specialist centres.

The role of cytology in the diagnosis of breast lesions is complementary with the clinical opinion of the examining surgeon and the mammographic appearances, the so-called 'triple approach'. When a woman attends a clinic with a self-discovered breast lump, or with a mammographic abnormality from a screening programme, she will be examined by a surgeon who will note the features of the lesion (palpable/impalpable, fixed/mobile, tender/non-tender, etc.) and make an assessment as to the likelihood of the lesion being benign or malignant. This is often formalised as a score between 1 and 5 with 1 being normal, 2 abnormal but benign, 3 suspicious probably benign, 4 suspicious probably malignant and 5 definitely malignant. If the woman has not already had a mammogram then this will be performed and a radiologist will make an assessment of the likelihood of a benign or malignant diagnosis and will usually express this by the same numerical score. The cytopathologist will view the FNAB and make a diagnosis, again expressed by the score and a text statement of the findings. The information from these three sources will be reviewed in a multi-disciplinary team meeting with the surgeon, radiologist and cytopathologist present, and a final integrated assessment of the likelihood of a benign or malignant process will be made. A malignant diagnosis of breast cancer is highly likely to lead to surgical treatment, such as removal of the whole breast (mastectomy) or part of it (wide local excision), so the whole diagnostic process must have a very high specificity with as few false positives as possible. There is really no acceptable rate of false positives since deforming surgery on a woman without breast cancer must be avoided at all costs. However if the specificity is set very high then the sensitivity may be lower and women with breast cancer may
not have their disease detected. For this reason there is often conscious, or 
unconscious, agreement between the surgeon, radiologist and cytopathologist that 
their tests, and scores, will have different ranges of sensitivity and specificity. Since 
mammography is used as a screening test it needs to have a relatively high sensitivity 
and clinical examination is often made with a high sensitivity in mind. This means 
that the cytodiagnosis of breast cancer must be carried out with a high specificity to 
reduce the number of false positives in the integrated diagnostic process. Again there 
is no real acceptable rate of false positive diagnosis but rates should certainly be less 
than 1% and preferably less than 0.1%. This requirement for high specificity is 
relatively unusual in the medical domain since it is usually sensitivity that is at a 
premium, e.g. in the microbiological diagnosis of bacterial meningitis, because 
usually medical treatments are fairly non-damaging, e.g. a course of antibiotic drugs, 
in comparison with the untreated disease.

**Study population**

692 consecutive adequate specimens of fine needle aspirates of breast lumps (FNAB) 
received at the Department of Pathology, Royal Hallamshire Hospital, Sheffield, 
during 1992 -1993. The final outcome of benign disease or malignancy was confirmed 
by open biopsy where this result was available. In benign aspirates with no 
subsequent open biopsy a benign outcome was assessed by clinical details on the 
request form, mammographic findings (where available) and by absence of further 
malignant specimens. A malignant outcome was confirmed by histology of open 
biopsy or clinical details where the primary treatment modality was chemotherapy or 
hormonal therapy.

**Input variables**

The eleven input features were the patient age in years and observations of the ten 
defined features given in table 1. All observations were made by a consultant 
pathologist with 10 years experience of reporting FNABs. All features were coded in 
binary format 0 = feature absent, 1 = feature present.
Table 1. The defined human observations used as input variables.

<table>
<thead>
<tr>
<th>Observed Feature</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cellular dyshesion</td>
<td>True if the majority of epithelial cells are dyshesive, false if the majority of epithelial cells are in cohesive groups</td>
</tr>
<tr>
<td>Intracytoplasmic lumina</td>
<td>True if intracytoplasmic lumina are present in some epithelial cells, false if absent</td>
</tr>
<tr>
<td>'Three-dimensionality' of epithelial cells clusters</td>
<td>True if some clusters of epithelial cells are not flat (more than two nuclei thick) and this is not due to artefactual folding, false if all clusters of epithelial cells are flat</td>
</tr>
<tr>
<td>Bipolar 'naked' nuclei</td>
<td>True if bipolar 'naked' nuclei are present, false if absent</td>
</tr>
<tr>
<td>Foamy macrophages</td>
<td>True if foamy macrophages are present, false if absent</td>
</tr>
<tr>
<td>Nucleoli</td>
<td>True if more than three easily-visible nucleoli are present in some epithelial cells, false if three or fewer easily-visible nucleoli in all epithelial cells</td>
</tr>
<tr>
<td>Nuclear pleomorphism</td>
<td>True if some epithelial cells have nuclear diameters twice that of other epithelial cell nuclei, false if no epithelial cell nuclei have diameters twice that of other epithelial cell nuclei</td>
</tr>
<tr>
<td>Nuclear size</td>
<td>True if some epithelial cell nuclei have diameters twice that of red blood cell diameters, false if all epithelial cell nuclei have diameters less than twice that of red blood cell diameters</td>
</tr>
<tr>
<td>Necrotic epithelial cells</td>
<td>True if necrotic epithelial cells are present, false if absent</td>
</tr>
<tr>
<td>Apocrine change</td>
<td>True if the majority of epithelial cell nuclei show apocrine change, false if apocrine change is not present in the majority of epithelial cells</td>
</tr>
</tbody>
</table>
Partitioning of the dataset
The data were randomly ordered in the order given in the spreadsheet. The data can be partitioned into a training set of the first 231, an optimisation/verification set of the next 231 and a test set of the final 230 cases. If a method is used which does not require an optimisation/verification set then the first 432 cases can be used as a training set.

Expert human performance on the dataset
After making the defined observations the human observer gave a categorical benign or malignant diagnosis without the use of a suspicious category. The appropriate metrics for the comparison are the sensitivity, specificity and predictive values of the tests; all with calculated 95% confidence intervals. These diagnoses gave the following performance:

Table 2. Human performance on the whole 692 item data set.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value with 95% confidence intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>82% (77-87)</td>
</tr>
<tr>
<td>Specificity</td>
<td>100%</td>
</tr>
<tr>
<td>Predictive value of a positive result</td>
<td>100%</td>
</tr>
<tr>
<td>Predictive value of a negative result</td>
<td>92% (89-94)</td>
</tr>
</tbody>
</table>

Our analyses of the data
We have used the data (not always all 692 cases) in logistic regression,(11) multilayer perceptron artificial neural networks, adaptive resonance theory mapping (ARTMAP) neural network(12-15) and a novel GCS system.(11)

A logistic equation was derived from the 432 case combined training set entering all variables together in a main effects only model and this was applied to the 230 case test set. The logistic regression was implemented using the Statistical Package for
Social Sciences (SPSS, http://www.spss.com/UK) running on a standard computer. The receiver operating characteristic (ROC) curve was calculated for the test set using standard methods (16).

Figure 1. ROC curve for logistic regression.

At a threshold of 0.5 this logistic equation gave the following results for the test set:

Table 3. Performance of logistic regression at a threshold of 0.5.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value with 95% confidence intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>94% (89-99)</td>
</tr>
<tr>
<td>Specificity</td>
<td>95% (90-97)</td>
</tr>
<tr>
<td>Predictive value of a positive result</td>
<td>87% (80-95)</td>
</tr>
<tr>
<td>Predictive value of a negative result</td>
<td>97% (95-99)</td>
</tr>
</tbody>
</table>
References


