Monograph:
Cross, S.S. and Harrison, R.F. (1998) The Department of Automatic Control and Systems Engineering research reports offer a forum for the research output of the academic staff and research students of the Department at the University of Sheffield. Papers are reviewed for quality and presentation by a departmental editor. However, the contents and opinions expressed remain the responsibility of the authors. Some papers in the series may have been subsequently published elsewhere and you are advised to cite the later published version in these instances. ing a Decision Support System for the Histopathalogical Diagnosis of Chronic Idiopathic Inflammatory Bowel Disease-Comparison of Radial Basis Function Neural Networks and Logistic Regression. UNSPECIFIED. ACSE Research Report 727 . Department of Automatic Control and Systems Engineering

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Developing a Decision Support System for the Histopathological Diagnosis of Chronic Idiopathic Inflammatory Bowel Disease – Comparison of Radial Basis Function Neural Networks and Logistic Regression

Dr Simon S Cross, *Dr Robert F Harrison
Department of Pathology
University of Sheffield Medical School
Beech Hill Road
Sheffield S10 2RX

*Department of Automatic Control and Systems Engineering
University of Sheffield
Mappin Street
Sheffield S1 3JD

Correspondence to :   Dr. S. S. Cross,
                             Department of Pathology
                             University of Sheffield Medical School
                             Beech Hill Road
                             Sheffield S10 2RX
                             UK
                             Tel: +44(0)114 2712683
                             Fax: +44(0)114 2780059
                             Email: s.s.cross@sheffield.ac.uk

1

Introduction:

The medical problem domain that is investigated in this study is the histopathological diagnosis of chronic idiopathic inflammatory bowel disease (CIIBD). CIIBD is a generic category that describes diseases of the bowel which are characterised by acute and chronic inflammation and which have no identified aetiological agent (such as an infective agent). The two major diseases within this category are ulcerative colitis (UC) and Crohn's disease. Both diseases are chronic conditions characterised by periods of relapse and remission and may produce life-threatening complications such as intestinal perforation, sepsis and carcinoma. Many conditions mimic the clinical symptoms of CIIBD (Farmer, 1990; Hamilton, 1987; Moxson et al. 1994; Shepherd, 1991; Shivananda et al. 1991; Surawicz et al. 1994) and thus histopathological examination of colorectal biopsies is important, both in confirming the diagnosis and in excluding other conditions, such as infective colitis (Surawicz et al. 1994; Nostrant et al. 1987; Surawicz and Belic, 1984) which may be made worse by steroid administration. In addition, distinction may be made between ulcerative colitis and Crohn's disease allowing appropriate treatment to be planned (Grobler et al. 1993; Handelsman et al. 1993; Hyman et al. 1991).

Samples for the histopathological diagnosis of CIIBD are taken from the colon at endoscopic examination. These small biopsies (2 mm in diameter) are embedded in paraffin wax and thin sections are stained with haematoxylin and eosin to be examined by light microscopy. The histopathological diagnosis is made subjectively by trained histopathologists. Histopathologists acquire their knowledge and decision-making processes from textbooks and from teaching by more experienced histopathologists, often using a double-headed microscope so teacher and pupil are viewing the same image. In Britain trainee histopathologists (who will be medically-qualified) are required to have 5 years postgraduate training in recognised laboratories before they can take the final examinations of the Royal College of Pathologists and be eligible to become consultant histopathologists. The diagnostic process in histopathology is poorly-understood but is believed to be a combination of pattern recognition and some form of heuristic logic (Underwood, 1987). The latter is not, however, formulated in any published form for the diagnosis of CIIBD.

The performance of histopathologists in the diagnosis of CIIBD has been investigated by a few published studies but most of these have been carried out in specialist centres for identified studies and so are likely to represent better performance than the overall standard. However these studies produce a sensitivity for the diagnosis of CIIBD in the range of 40% to 82% and a specificity for the diagnosis of CIIBD in the range of 73% to 98% (Frei and Morson, 1981; Thompson et al. 1985; Jenkins, 1988; Seldenrijk et al. 1991; Surawicz et al. 1994). There is thus scope for a decision-support system in the histopathological diagnosis of CIIBD to improve the sensitivity and PPV of fully-trained histopathologists and for use in the long training period required for histopathology novices. Here we describe the development of such a decision support system.

Methods:

Study population

The study population was drawn from large bowel endoscopic biopsies reported in the Department of Histopathology, Royal Hallamshire Hospital, Sheffield between 1990 and 1995 (inclusive). Biopsies originating in diverted bowel, rectal stumps or pouches were excluded, as were those with a diagnosis of neoplasm. The diagnosis was confirmed by the finding of typical endoscopy appearances seen on video photographs in the clinical notes, subsequent bowel resection, pattern of disease on radiological investigation or microbiological culture results. In cases without confirmation by subsequent resection specimens this final diagnostic outcome was made with review of the patient's case notes. This produced a set of 487 endoscopic biopsies with outcomes of normality in 84 cases, Crohn's disease in 120 cases and ulcerative colitis in 283 cases. The biopsies were a mixed population of single distal biopsies and colonoscopic series from initial presentation and follow-up of disease. The biopsies were examined (blind to all clinical details) by a single experienced observer (SSC) using a computer interface which implements the BSG Guidelines for the Assessment (Jenkins et al. 1997) with digitised images representing examples of each histopathological feature (Cross et al. 1997). Some of the features are dichotomous variables, e.g. the presence or absence of mucosal granulomas, whilst others are ordinal categories, e.g. mucin depletion classified into none, mild, moderate or severe. The observed features and their coding are given in table 1. Observation was spread over a period of 9 months with no more than 30 biopsies observed in a single day.

Neural network and logistic regression analysis

The neural network adopted for this study was based on the radial basis function network (RBFN) (Moody & Darken, 1989); Broomhead & Lowe, 1988), using a SoftMax (Bridle, 1990) output layer to ensure that outputs may be interpreted as probabilities, and Gaussian radial basis functions. In one-from-many classification problems such as this one, the natural performance measure is not the more usual mean (or total) square error, but rather the cross-entropy function (Bridle, 1990). This matches the SoftMax outputs and leads to a nice Bayesian interpretation (Bishop, 1995). The gradient descent learning rule derived from these considerations is identical in form to the celebrated Widrow-Hoff (Widrow & Hoff, 1960) or delta rule (Rumelhart & McClelland, 1986). Here we use batch training. Kernel centres are selected simply by choosing a number of training vectors arbitrarily, and the kernel widths are fixed for each experiment. Finally a weight decay term is introduced to regularise the learning procedure (i.e. to prevent overspecialisation to the training data) (Bishop, 1995; Ripley, 1996).

The necessary feedforward equations are:

$$y_i = \frac{\exp(net_i)}{\sum_{j=1}^{N} \exp(net_j)}; \quad net_j = \sum_{k=1}^{R} w_{jk} z_k; \quad z_k = \exp\left(\frac{\|\mathbf{x} - \mathbf{c}_k\|^2}{\sigma^2}\right)$$

where i ranges over the number, N, of classes, R is the number of radial basis functions, w_{jk} the weight between the kth radial basis function and the jth output unit, $\mathbf{x}$ the input vector, $\mathbf{c}_k$ the kth centre vector, and σ the (uniform) radial basis function width. $\|\mathbf{v}\|$ denotes the (Euclidean) length of $\mathbf{v}$. The learning rule is given by:

$$\mathbf{w}_j(t+1) = (1-\rho)\mathbf{w}_j(t) + \lambda \sum (t_j - y_j)\mathbf{z}$$

where $\mathbf{w}_j$ is the weight vector associated with the jth output node, $\mathbf{z}$ the vector of radial basis function layer activations, ρ the weight decay (regularisation) parameter $(0 < \rho \ll 1)$, λ the learning rate, t_j the target output for the jth node, and the summation takes place over all patterns in the training set.

The input data were normalised to the interval (-1,1). This corresponds only to a shift in origin for the binary features and can be justified for the categorical features on the grounds that they are ordinal and that increasing value corresponds to an increasing effect. Identical input data were used in the logistic regression trials.

The Crohn's disease and ulcerative colitis cases were combined into a single generic category of chronic idiopathic inflammatory bowel disease to be distinguished from the normal cases. The ratio of cases was skewed markedly towards the CIIBD cases (403 CIIBD, 84 normal; mainly due to the difficulty of confirming a "normal" outcome) so a bootstrap method was used to ascertain the best configuration. 60 cases from each category were selected randomly for the training set and the remaining 24 normal cases and 24 randomly selected CIIBD cases were used as the test set, i.e. both the training and the evaluation stages were bootstrapped. This selection process was repeated 1000 times for the neural network and logistic regression techniques.

Our chosen measure of performance is the area beneath the receiver operating characteristic (ROC) curve (Hanley & McNeil, 1982). This allows the performance of different classifiers to be compared independent of the choice of decision threshold, thereby giving a single figure of merit for each classifier (Hanley & McNeil, 1983).

In comparing two classifiers developed using the same data samples it is necessary to take into account any correlation between their outputs. Not surprisingly, this can often be high and can strongly effect the test of significance (Hanley & McNeil, 1983). Here, owing to the bootstrap approach, it is not feasible to calculate the required correlation coefficients. Instead we take a parametric approach and find the value of correlation coefficient (ρ_{sig}) above which differences in ROC area become significant (at the 95% level).

Results:
The results are summarised in tables 2 and 3. The best-performing neural network with the 7 selected input variables had 20 radial basis functions of width 1.0, regularisation parameter of 0.001 and learning rate of 0.0001. The best-performing neural network using all 24 input variables has 20 radial basis functions of width 0.5, regularisation parameter of 0.0005 and learning rate of 0.0001.

Discussion:
The problem domain of the histopathological diagnosis of CIIBD is interesting because it is not entirely clear which parameters (sensitivity, specificity etc.) should be optimised in a decision support system. Optimising sensitivity over specificity will produce more false positives so some patients will be labelled as having CIIBD when they do not and will be subject to unnecessary long-term follow-up including regular colonoscopy - which is time-consuming, costly and uncomfortable procedure. Optimising specificity may lead to cases of CIIBD being mislabelled as normal and so not receiving the long-term follow-up which is recommended to detect epithelial dysplasia in the colon and so prevent the development of carcinoma. However since CIIBD is characterised by episodes of disease activity and remission it is likely that cases misdiagnosed as normal would re-present with symptoms at some time in the future. With such uncertainty about which parameter to optimise our choice of selecting an "optimal" threshold at which sensitivity and specificity are equal seems a reasonable choice. This highlights the advantage of using a threshold-free performance measure, such as the area under the ROC curve, for comparative purposes.

At our chosen thresholds the performance of the radial basis function neural networks gave a sensitivity better than 4 published studies of human performance (Frei and Morson, 1981; Thompson et al. 1985; Jenkins, 1988; Seldenrijk et al. 1991) and 5% less than the other published study (Surawicz et al. 1994). The specificity was better than two of the published studies (Thompson et al. 1985; Jenkins, 1988) but 11-23% less than the other three studies (Frei and Morson, 1981; Seldenrijk et al. 1991; Surawicz et al. 1994) although only one of those studies (Frei and Morson, 1981) is an audit of routine laboratory performance and originates from a centre specialising in gastrointestinal disease. It would appear that the histopathologists in these studies have, consciously or unconsciously, set a higher threshold for the diagnosis of CIIBD at the expense of lower sensitivity. It would be interesting to investigate whether a decision support system developed using a decision-making engine similar to this study and used by trained histopathologists would produce an increase in sensitivity without loss of specificity. The overall performance of our system appears to be good enough to be used as an aid for trainee histopathologists but it is dependent on reliable identification of the observed microscopic features and this aspect requires further study.

Both logistic regression and the neural networks gave better performance with 7 selected inputs (variables 3, 4, 5, 6, 7, 17 and 22 in table 1) rather than all 24 available inputs (comparison shown in tables 2 and 3). The 7 selected inputs were those highlighted by the guidelines published under the auspices of the British Society for Gastroenterology (Jenkins et al. 1997) as being important in distinguishing CIIBD from normality after an evidence-based meta-analysis of published studies. This study would appear to provide further evidence that these variables are important in these category assignments and the other variables may be acting as 'noise' resulting in degraded performance.

The performance of the neural networks will only be significantly better than logistic regression when the correlation between the outputs of these systems is close to one (i.e. both systems are making the same correct and incorrect predictions). Owing to the limitations of the available software used in the study it has not been possible to calculate the required correlation coefficients to determine if the apparently better performance of the neural networks compared with logistic regression is statistically significant.

References:
Bishop, C.M. (1995) Neural Networks for Pattern Recognition. Oxford: Clarendon Press.

Bridle, J.S. (1990) Training stochastic model recognition algorithms as networks can lead to maximum mutual information estimation of parameters. In Touretzky, D., S. (Ed.), Advances in Neural Information Processing 2. San Mateo: Morgan Kauffman.

Broomhead, D.S. & Lowe, D. (1988) Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2, 321-355.

Cross, S.S., Dube, A.K., Underwood, J.C.E., Pialle, H., Sharkey, A. and Jenkins, D. (1997) A visual image-based data entry system for reporting non-neoplastic endoscopic colorectal biopsies. *J.Pathol.* **181**, 17A(Abstract)

Farmer, R.G. (1990) Infectious causes of diarrhoea in the differential diagnosis of inflammatory bowel disease. *Med.Clin.N.Am.* **74**, 29-38.

Frei, J.V. and Morson, B.C. (1981) Medical audit of rectal biopsy diagnosis of inflammatory bowel disease. *J.Clin.Pathol.* **35**, 341-344.

Grobler, S.P., Hosie, K.B., Affie, E., Thompson, H. and Keighley, M.R.B. (1993) Outcome of restorative protocolectomy when the diagnosis is suggestive of Crohn's disease. *Gut* **34**, 1384-1388.

Hamilton, S.R. (1987) The differential diagnosis of idiopathic inflammatory bowel disease by colorectal biopsy. *Int.J.Colorect.Dis.* **2**, 113-117.

Handelsman, J.C., Gottlieb, L.M. and Hamilton, S.R. (1993) Crohn's disease as a contraindication to Kock pouch (continent ileostomy). *Dis.Col.Rectum* **36**, 840-843.

Hanley, J.A. & McNeil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29-36.

Hanley, J.A. & McNeil, B.J. (1983) A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148, 839-843.

Hyman, N.H., Fazio, V.W., Tuckson, W.B. and Lavary, I.C. (1991) Consequences of ileal pouch-anal anastomosis for Crohn's colitis. *Dis.Col.Rectum* **34**, 653-657.

Jenkins, D. (1988) Computing and histopathology of intestinal inflammation. In: Vicery, F.R., (Ed.) *Computers in Gastroenterology*, pp. 193-204. London: Springer-Verlag]

Jenkins, D., Balsitis, M., Gallivan, S., Dixon, M.F., Gilmour, H.M., Shepherd, N.A., Theodossi, A. and Williams, G.T. (1997) Guidelines for the initial biopsy diagnosis of suspected chronic idiopathic inflammatory bowel disease. The British Society of Gastroenterology Initiative. *J.Clin.Pathol.* **50**, 93-105.

Moody, J. & Darken, C. (1989) Fast learning in networks of locally-tuned processing units. *Neural Computation*, 1, 281-294.

Moxson, C.J., Klein, H.D. and Rubin, W. (1994) Atypical forms of inflammatory bowel disease. *Med.Clin.N.Am.* **78**, 1259-1273.

Nostrant, T.T., Kumar, N.B. and Appelman, H.D. (1987) Histopathology differentiates acute self-limited colitis from ulcerative colitis. *Gastroenterology* **92**, 318-328.

Ripley, B.D. (1996) Pattern Recognition and Neural Networks. Cambridge, UK: Cambridge University Press.

Rumelhart, D.E. & McClelland, J. (1986) Parallel Distributed Processing. Cambridge MA: MIT Press.

Seldenrijk, C.A., Morson, B.C., Meuwissen, S.G., Schipper, N.W., Lindeman, J. and Meijer, C.J. (1991) Histopathological evaluation of colonic mucosal biopsy specimens in chronic inflammatory bowel disease: diagnostic implications. *Gut* **32**, 1514-1520.

Shepherd, N.A. (1991) Pathological mimics of chronic inflammatory bowel disease. *J.Clin.Pathol.* **44**, 726-733.

Shivananda, S., Hordijk, M.L., Ten Kate, F.J.W., Probet, S.J. and Mayberry, J.F. (1991) Differential diagnosis of inflammatory bowel disease: a comparison of various diagnostic classifications. *Scand.J.Gastroenterol.* **26**, 167-173.

Surawicz, C.M. and Belic, L. (1984) Rectal biopsy helps to distinguish acute self-limited colitis from idiopathic inflammatory bowel disease. *Gastroenterology* **86**, 104-113.

Surawicz, C.M., Haggitt, R.C., Husseman, M. and McFarland, L.V. (1994) Mucosal biopsy diagnosis of colitis: acute self-limited colitis and idiopathic inflammatory bowel disease. *Gastroenterology* **107**, 755-763.

Thompson, E.M., Price, A.B., Altman, D.G., Sowter, C. and Slavin, G. (1985) Quantitation in inflammatory bowel disease using computerised interactive image analysis. *J.Clin.Pathol.* **38**, 631-638.

Underwood, J.C.E. (1987) *Introduction to Biopsy Interpretation and Surgical Pathology*, 2 edn. London: Springer-Verlag.

Widrow, B. & Hoff, M. (1960) Adaptive switching circuits. Proceedings of the 1960 IRE WESCON Convention Part 4. (pp.96-104)

Table 1. The observed features (input variables) and their coding.

Variable No.	Input Variable	Type	Values
1	Age	real number	16 to 86
2	Female sex	binary	0,1
3	Mucosal surface	ordinal categorical	0,1,2,3
4	Crypt architecture	ordinal categorical	0,1,2,3
5	Crypt profiles	real number	2 to 8
6	Increased lamina propria cellularity	binary	0,1
7	Patchy increased lamina propria cellularity	binary	0,1
8	Increased lymphoid lamina propria cellularity	binary	0,1
9	Diffusely increased lamina propria cellularity	binary	0,1
10	Diffuse transmucosal inflammation	binary	0,1
11	Cryptitis - extent	ordinal categorical	0,1,2,3
12	Cryptitis - polymorphs	ordinal catergorical	0,1,2,3
13	Crypt abscesses - extent	ordinal categorical	0,1,2,3
14	Crypt abscesses - polymorphs	ordinal categorical	0,1,2,3
15	Focal lamina propria polymorphs	binary	0,1
16	Diffuse lamina propria polymorphs	binary	0,1
17	Epithelial flattening	binary	0,1
18	Epithelial degeneration	binary	0,1
19	Epithelial erosion	binary	0,1
20	Mucin depletion	ordinal categorical	0,1,2,3
21	Intraepithelial lymphocytes	binary	0,1
22	Lamina propria granulomas	binary	0,1
23	Submucosal granulomas	binary	0,1
24	Basal histiocytes	binary	0,1

7

Table 2. Areas under the ROC curves for logistic regression (LR) and radial basis function neural networks (NN) for 7 and 24 input variables.

	LR area (SE)	NN area (SE)	ρ_{sig}
7 inputs	.773 (.066)	.807 (.061)	.97
24 inputs	.721 (.072)	.758 (.065)	.97
ρ_{sig}	.93	.93	

Table 3. The sensitivity and specificity for LR and NN operating at their "optimal" threshold (when sensitivity and specificity are equal).

	LR Sensitivity and specificity (%)	NN Sensitivity and specificity (%)
7 inputs	74	77
24 inputs	70	68