



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/82441/>

Monograph:

Cross, S.S. and Harrison, R.F. (1998) Making the Distinction Between Crohn's Disease and Ulcerative Colitis by Histopathological Examination: A Comparison of Human Performance, Logistic Regression and Adaptive Resonance Theory Mapping Neural Networks (ARTMAP). Research Report. ACSE Research Report 726 . Department of Automatic Control and Systems Engineering

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

X

Making the Distinction Between Crohn's Disease and Ulcerative Colitis by Histopathological Examination: A Comparison of Human Performance, Logistic Regression and Adaptive Resonance Theory Mapping Neural Networks (ARTMAP)

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
Research Report Number 726

Abstract:

280 cases of inflammatory bowel disease were examined histopathologically and defined features were observed and recorded using a graphical user interface with digitised reference images. The outcome of each case was determined independent of the histopathological report giving 75 cases of Crohn's disease and 105 cases of ulcerative colitis. The cases were randomised and split into training and test sets, each of 140 cases. All 23 observed features were used as input data for logistic regression and adaptive resonance theory mapping neural networks (ARTMAP). The ARTMAPs were used singly or as a voting cohort of 11 networks, the majority and unanimous decisions of the cohort were analysed separately. The vigilance parameter for the ARTMAPs was varied from high (0.9) to lower (0.5) to vary the number of clusters in ART_a. The best results were produced by logistic regression and the 11 high vigilance ARTMAPs with a sensitivity for Crohn's disease of 60% and a specificity of 80%. This was a significant improvement on the original human diagnoses that gave a sensitivity of 25% and specificity of 60%. Either of these technologies could form the basis of a decision support system in this domain. (193)

Keywords: Crohn's disease, ulcerative colitis, inflammatory bowel disease, logistic regression, adaptive resonance mapping neural networks, ARTMAP

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



200446983



Making the Distinction Between Crohn's Disease and Ulcerative Colitis by Histopathological Examination: A Comparison of Human Performance, Logistic Regression and Adaptive Resonance Theory Neural Network Mapping (ARTMAP)

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

Introduction:

Histopathology is usually recognised as the 'gold standard' for the 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. The two major diagnostic decisions to be made in this area are: 1. Does the subject have CIIBD or does she/he have some other cause for the inflammation in their bowel? 2. If the subject does have CIIBD is it Crohn's disease or ulcerative colitis? We have already investigated the first question using logistic regression and radial basis function neural networks (Cross & Harrison, 1998). In this study we use logistic regression and adaptive resonance theory mapping neural networks (ARTMAP) to classify cases of CIIBD into Crohn's disease or ulcerative colitis and compare this with the original human diagnosis.

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 a trainee histopathologist (who will be medically-qualified) is required to have 5 years postgraduate training in recognised laboratories before she/he can take the final examinations of the Royal College of Pathologists and be eligible to become a consultant histopathologist. 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) but the latter is not 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 Crohn's disease or ulcerative colitis in the range of 40% to 82% and a specificity for these diagnoses 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.

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 280 endoscopic biopsies with outcomes of Crohn's disease in 75 cases and ulcerative colitis in 205 cases. The biopsies were a mixed population of single distal biopsies and colonoscopic series from initial presentation and follow-up of disease, all the biopsies showed active inflammation as evidenced by the presence of neutrophil polymorphs. 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.

Human Performance

The initial reports of the biopsies were examined and the original histopathological diagnosis was extracted and classified into one of the categories specified in the BSG Guidelines (Jenkins et al. 1997).

Logistic Regression and Neural Network Analysis

The 280 biopsies were randomised into two sets of 140. One set was used for training the neural network or deriving a logistic equation and the other set was used as an independent test set. All 23 input variables were used in all analyses. The neural networks used were a custom implementation of the ARTMAP architecture (Downs et al. 1996). These networks were used in two different configurations - singly and in voting cohorts. The single ARTMAPs were trained on the 140 set training set and tested on the second set and the best performances were selected. In the voting cohort 11 single ARTMAPs were trained and the decisions were made by taking the majority vote of these networks. In many cases the 11 ARTMAPs produced a unanimous decision and the performance in such circumstances was analysed separately. The vigilance parameter of the ARTMAPs was set very high (0.9) during training in order to maximise classification performance. Vigilance was relaxed to 0.6 for predictions on the test set to ensure that all cases were matched to an existing category cluster node. In a second training run the vigilance parameter was lowered to 0.5 to reduce the number of clusters in ART_a.

Results:

The results are summarised in table 2. The trained ARTMAPs contained 76-85 nodes in ART_a with a training vigilance parameter of 0.9 and 14-24 nodes with a training vigilance parameter of 0.5.

Discussion:

Table 2 shows that any classifier derived from the human observations produced an improved performance when compared with the original human performance. The improvements were substantial with an increase in sensitivity for Crohn's disease of 35% and increase in specificity of 23%. The best overall performance was given by either logistic regression or 11 voting ARTMAPs (vigilance parameter 0.9), these two methods had virtually identical performance parameters. The most robust specificity (and thus predictive value of a positive test) was given by the unanimous decisions of the 11 ARTMAPs that covered 104 of the 140 cases in the test set. The number of ART_a clusters was reduced by about two-thirds by reducing the vigilance parameter during training (and thus giving a more generalised solution) but there was some degradation in performance. An improved technique might have been a training phase followed by pruning of low confidence clusters and then testing but there were insufficient cases to provide 3 adequate sets.

The original human performance has been derived from the original reports that were made by a large number of different histopathologists. All reporting was supervised by consultant histopathologists but the initial report, before checking, may have been written by trainee pathologists with varying levels of experience. The reports were not constrained by a dichotomous ulcerative colitis/Crohn's disease output and many cases were reported as inflammatory bowel disease of indeterminate histological type. The observations used by the statistical classifiers were made by a single consultant histopathologist with 14 years experience of histopathology and a specialist interest in gastrointestinal

pathology. All these factors may have contributed to the increases in performance shown by the statistical classifiers but the increases are large which suggest that there is also a positive contribution from the statistical technology. In this study there are a large number of input variables (23) and it is likely that logistic regression or ARTMAPs are a more efficient method of weighting these factors when making a decision than an unassisted human brain which is more efficient at pattern recognition in histopathological diagnosis.

In this study there was an imbalance in representation of Crohn's disease and ulcerative colitis (75 v. 205 cases) but this reflects the *a priori* incidence of these disease and so it is more realistic to use these proportions in the training and test sets of any possible decision support system. An uncertainty in this domain is the selection of parameters to optimise in any such system. The treatment of Crohn's disease and ulcerative colitis diverges most markedly when surgical intervention is required. In ulcerative colitis complete resection of the colon will cure the disease because it is only present in the colon. An ileoanal pouch can be constructed after resection of the colon to obviate the need for a permanent cutaneous stoma. In Crohn's disease the inflammatory process may affect any part of the intestine and any surgical interventions try to conserve as much bowel as possible. Ileoanal pouches are contraindicated in Crohn's disease because of the risk of fistula formation. This information suggests that the sensitivity for Crohn's disease should be optimised and this is certainly important to avoid surgical complications. However if this increased sensitivity results in the misclassification of some subjects with ulcerative colitis as Crohn's disease then these subjects may have surgical treatment leaving them with a permanent cutaneous stoma when they could have had an ileoanal pouch. It thus appears that optimisation of overall accuracy is probably the best procedure in this domain.

The choice of which method to implement as a decision support system will be more a choice of implementation since logistic regression and majority voting ARTMAPs given virtually identical results. A logistic equation would appear to be easier to implement and could easily be produced using generic spreadsheet software (e.g. Microsoft Excel) with a user-friendly front end. An extended study with a larger number of cases would probably be required before the production and testing of such a system.

References:

- Cross, S.S., Dube, A.K., Underwood, J.C.E., Piale, 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)
- Cross SS, Harrison RF. 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. NEURAP'98, IUSPM, University of Aix-Marseille.
- Downs, J., Harrison, R.F., Kennedy, R.L. and Cross, S.S. (1996) Application of the fuzzy ARTMAP neural network model to medical pattern classification tasks. *Artificial Intelligence in Medicine* **8**, 403-428.
- Frei, J.V. and Morson, B.C. (1981) Medical audit of rectal biopsy diagnosis of inflammatory bowel disease. *J.Clin.Pathol.* **35**, 341-344.
- 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.
- 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.
- 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.



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 categorical	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

Table 2 - Results of the various classifiers expressed in terms of Crohn's disease diagnosis. 95% confidence intervals are given in parentheses.

Classifier	Sensitivity for Crohn's disease	Specificity for Crohn's disease	Predictive value of +ve test for Crohn's disease	Predictive value of -ve test for Crohn's disease
Original human performance ¹	25% (15-35)	60% (53-66)	19% (11-26)	69% (27-75)
Logistic regression ²	60% (45-74)	82% (75-90)	60% (45-74)	82% (75-90)
Single ARTMAP ²	50% (35-65)	92% (86-97)	72% (56-89)	81% (74-88)
11 voting ARTMAPs ² (vigilance 0.9)	60% (45-74)	83% (75-90)	60% (45-74)	83% (75-90)
Unanimous decisions of 11 ARTMAPs ³ (vigilance 0.9)	48% (28-68)	100%	100%	86% (80-93)
11 voting ARTMAPs ² (vigilance 0.5)	45% (30-60)	82% (74-89)	51% (35-67)	78% (70-86)

1. On all 280 cases
2. On the 140 case test set
3. On the 108 cases from the 140 case test set which had unanimous decisions