Pre-publication draft of a paper which appeared in Sarmeinto, C. (Ed.), *Proceedings of the IJCAI Workshop W.15: Representing Knowledge in Medical Decision Support Systems*, pages 8.1-8.8.

The application of machine learning to the diagnosis of glomerular disease

Geoffrey I. Webb and John W. M. Agar

Deakin University and Geelong Hospital, Geelong, Victoria, Australia.

Abstract

A pilot study has applied the DLG machine learning algorithm to create expert systems for the assessment and interpretation of clinical and laboratory data in glomerular disease. Despite the limited size of the data-set and major deficiencies in the information recorded therein, for one of the conditions examined in this study, microscopic polyarteritis, a consistent diagnostic accuracy of 100% was obtained. With expansion of the data base, it is possible that techniques will be derived that provide accurate non-invasive diagnosis of some cases of glomerular disease, thus obviating the need for renal biopsy. Success in this project will result in significant reductions in both the cost and the morbidity associated with the investigation of glomerular disease.

1 Introduction

The central diagnostic procedure for patients with suspected glomerular disease involves the evaluation of percutaneous needle biopsy specimens by histochemical, immunological and electron-microscopic analysis. The results of these evaluations lead to a histological classification which serves as the diagnosis. This classification is in turn used to select therapeutic regimens and to identify prognostic implications.

During initial assessment, prior to renal biopsy, clinicians routinely record a full history and examination and undertake a wide range of laboratory investigations of blood and urine to document renal function, to assess the significance and severity of haematuria and proteinuria and to assess underlying immunological mechanisms.

Although these investigations provide the clinician with suggestions as to the condition which a patient is suffering, there are no accepted techniques for deriving a definitive diagnosis there from. As a result, renal biopsy is performed for the majority of patients in order to accurately classify glomerular disease. This reliance on renal biopsy for final glomerular diagnosis has the undesirable features of an identifiable morbidity and mortality as well as the costs associated with hospital admission and histological evaluation of tissue samples.

In the light of these factors, it would be desirable if techniques for reliable diagnosis of glomerular disease without recourse to renal biopsy could be developed. Our investigations suggest that it might be possible to develop such techniques through the application of machine learning. To evaluate the feasibility of such a large scale project, we have undertaken a pilot study in which we applied DLG (Webb, l991a), a variant of the Aq machine learning algorithm (Michalski, 1984), to create expert systems for classifying pre-biopsy clinical and laboratory data. The training sets for this study come from a retrospective data-base of 284 consecutive patients who have undergone diagnostic renal biopsy (Agar & Webb, in preparation).

The results of this study lead us to believe that the application of machine learning to a sufficiently broad data-set is likely to result in the development of non-intrusive diagnostic techniques which are sufficiently accurate to eliminate, in some cases, the need for biopsy, with its associated cost/morbidity implications.

2 Background

Previous attempts to develop techniques for accurate non-surgical diagnostic classification of glomerular disease have been generally unsuccessful (Baak & Wehner, 1983; Urakabe, Orita, Shirai, et al., 1975). One possible exception is a study that used discriminant analysis to develop a set of discriminant functions from a sample of 138 patients with primary glomerular disease (Tomura, Tsutani, Sakuma, et a!., 1985). When applied to the sample from which it was developed, this set of discriminant functions identified the correct diagnosis in 76.4% of cases. However, a study employing a sample of several thousand patients was unable to replicate this result (personal communication, Linda Sharples and Adrian Morley on behalf of the MRC Treatment of Glomerulonephritis Group, August, 1990). This suggests that the discriminant functions developed do not generalize beyond the initial sample. Further, the sample contained only seven distinct histological categories and the discriminant functions that were induced relied upon other histological categories being excluded before application. This seriously reduces their potential utility as practical diagnostic criteria.

There is reason to believe, however, that machine learning may succeed in uncovering non-surgical diagnostic techniques where quantitative analysis techniques have failed. The DLG machine learning algorithm is based on formal logic as opposed to the statistical basis of most conventional medical research. Statistical techniques have difficulty detecting non-linear relationships between independent variables (clinical symptoms, physical examination and investigation results) and a dependent variable (diagnosis). Discriminant analysis, for example, may produce a measure for each diagnosis that is the sum of a number of measures each based on a single independent variable. For instance, it can detect that both raised blood pressure and high creatinine are independently critical factors in diagnosis. However, it cannot routinely detect that, for example, raised blood pressure is a critical factor in diagnosis if and only if the patient's creatinine is high. This problem is compounded when multiple variables interact (for example, blood pressure is only relevant if creatinine is high and urinary protein is raised or oedema is present).

Interactions between independent variables cause no such difficulties for the qualitative formal logic based techniques employed in the DLG algorithm. This makes it ideal for analysis of medical data where complex interactions between independent variables are common.

On account of these considerations, a pilot study has been undertaken to evaluate the feasibility of a large scale investigation of renal diagnosis using the DLG algorithm for analysis.

Such a project would differ from previous applications of machine learning to medical domains in that although definitive diagnosis is possible (by analysis of biopsy specimens), reliable diagnosis techniques do not yet exist for the independent variables to be examined. Some previous studies (for example, Quinlan, 1987 and Quinlan, Compton, Horn, et al., 1986) have induced diagnostic criteria with outstanding diagnostic accuracy (in excess of 99%) from data for which techniques already exist for definitive diagnosis from the independent variables. Other previous studies (for example, Michalski, Mozetic, Hong, & Lavrac, 1986) have induced diagnostic criteria which approach the accuracy of expert human diagnosis in domains for which definitive diagnostic techniques do not exist.

This study is also distinguished from previous machine learning studies by the large number of values for the dependent variable - in the pilot study, 31 diagnoses. Very few machine learning studies are detailed in the literature in which the dependent variable has more than five values.

3 Method

The pilot study utilised a data-base containing 284 consecutive case histories, each comprising a diagnosis and 38 variables recording clinical, biochemical and microbiological information (Table 1). The recorded diagnoses were derived by histochemical, immunological and electron - microscopic analysis of percutaneous renal biopsy specimens. All patients treated by a single clinician for which a renal biopsy was performed between January 1 1979 and July 31 1989 were

included. There were no exclusions on the basis of histological category or other grounds. Referral bias is minimal due to the use of a defined population base served by a single hospital and one nephrologist. The sample obtained is believed to be highly representative of glomerular disease in the wider Australian population. All histological assessment (Table 2) has been performed by a single department of renal histology (Prince Henry's Hospital, Melbourne), where histological diagnosis was established by the combination of standard light microscopy techniques, electron microscopy, immunofluorescence (1979 - 1986) and immunoperoxidase (1986 - 1989.)

The data-base is deficient, however, in that it was not created with the current study in mind, and thus does not include many variables that are believed to be relevant, such as family history, medication and drug ingestion, body weight and height, glomerular haematuria, red cell or other cellular casts and hepatitis status.

The DLG algorithm in the form of the Einstein[1] induction system installed on a Solboume 5/602 computer was applied to the data-base to create diagnostic expert systems.

A major objective of the study was to obtain an indication of the ability to derive rules from such data that can reliably diagnose previously unsighted cases. Due to the likelihood with such a small data-base of any single random distribution into training and evaluation sets giving unrepresentative results, 1000 such random distributions were performed. In each distribution, every case in the data set was given an 80% chance of being assigned to the training set. If it was not assigned to the training set then it was assigned to the evaluation set. The DLG algorithm was then applied to that training set and the resulting rules were evaluated against the associated evaluation set.

This process was also repeated 100 times for each diagnosis represented by more than 15 cases. In these latter analyses the derived expert system in each case examined the data and determined whether that single histological diagnosis applied.

Clinical data	**Laboratory data**	
Age	Urinary RBC Count	Haemoglobin
Sex	Urinary Total Protein	White Cell Count
Diagnosis (biopsy confirmed)	Creatinine Clearance	Platelets
Systolic Blood Pressure	Serum Creatinine	ESR
Diastolic Blood Pressure	Blood Urea	Antinuclear Factor
Oedema	Serum Total Protein	Anti DNA Antibody
Headache	Serum Albumin	Anti GBM Antibody
Dyspnoea	Serum IgG	
Haemoptysis	Serum IgA	
Arthralgia	Serum IgM	
Rash	C_3 Level	
Loin Pain	C_4 Level	
Upper Respiratory Infection	ASO Titre	
Anorexia	Cholesterol	
Weight Loss	Triglycerides	
Bowel Disturbance	Blood Glucose	

Table 1 Clinical and Laboratory Variables

[1] *Einstein is a trademark of UniLink Ltd, Deakin University, Geelong, Australia.*

Diagnosis	No of cases
Immunoglobulin A nephropathy	62
Membranous nephropathy	26
Mesangial proliferative glomerulonephritis	24
Focal glomerulosclerosis	20
Minimal lesion nephrotic syndrome	20
Minor changes	19
Microscopic polyarteritis	17
Lupus nephritis	16
Post infectious glomerulonephritis	13
Henoch Schonlein nephritis	9
Membranoproliferative glomerulonephritis	7
Anti GBM nephritis	7
Interstitial nephritis	6
Acute tubular necrosis	6
Diabetic nephropathy	5
Thin membrane disease	5
Alports syndrome	4
Amyloidosis	3
Ischaemic changes	2
Wegeners granulomatosis	2
Haemolytic uraemic syndrome	1
Idiopathic crescentic nephritis (RPGN)	1
Immunotactoid glomerulopathy	1
Leptospirosis	1
Multiple myeloma	1
Normal	1
Nephrosclerosis	1
Secondary oxalosis	1
Pre-eclampsia	1
Systemic sclerosis	1
Transplant nephropathy	1

Table 2 Renal Biopsy Diagnoses

4 Results

For the first series of analyses, the 1000 runs where expert systems for diagnosing between all 31 conditions were derived, the average accuracy obtained when the expert systems were applied to the training sets was 51.05%. That is, the 51.05% of the diagnoses reached by the expert systems derived from randomly selected training sets when applied to the remaining data were correct.

The results of the second series of analyses, when expert systems for diagnosing individual conditions were developed, are presented in Table 3. In this table, accuracy of positive diagnosis represents the percentage of cases of a particular histological category which were correctly diagnosed. Negative diagnosis represents the percentage of cases not belonging to a particular histological category which were correctly identified as not belonging to that category. Total accuracy represents the percentage of diagnoses made (either positive or negative) that were correct.

Figure 1 shows a set of rules for diagnosing microscopic polyarteritis developed by the Einstein system during this study.

5 Discussion

Percutaneous renal biopsy is currently essential for an accurate diagnosis of glomerular disease due to the absence of reliable non-invasive diagnostic tests. However, this surgical procedure is associated with significant cost implications and a small though identifiable patient morbidity and mortality. As discussed above, previous attempts to develop non-invasive diagnostic tests, in particular through statistical analysis, have been unsuccessful, failing to discover diagnostic criteria of sufficient accuracy for general clinical application.

In this context, we have used machine learning to develop rules for glomerular diagnosis. 284 consecutive patients presenting with suspected glomerular disease were assessed. All of these patients were diagnosed by histological assessment, which is accepted as the "gold standard" for glomerular diagnosis. 31 distinct conditions were present in our data-base. Only 44 clinical and laboratory variables were recorded omitting many which we believe are relevant.

Despite the small size and limited scope of the data-base and the large number of conditions represented, some successful diagnostic rules were generated.

Two different types of expert system were developed. The first type analysed a case and diagnosed which of the 31 conditions included in the study was present. The other type of expert system analysed a case and determined whether or not one particular condition was present.

Diagnosis	Accuracy of Positive Diagnosis (%)	Accuracy of Negative Diagnosis (%)	Total Accuracy (%)
Microscopic polyarteritis	100.00	95.34	95.37
Minimal lesion nephrotic syndrome	93.88	96.53	96.50
Immunoglobulin A nephropathy	75.11	82.12	81.26
Minor changes	63.64	93.73	93.66
Lupus nephritis	54.55	96.37	96.27
Focal glomerulosclerosis	36.36	92.20	92.06
Mesangial proliferative glomerulonephritis	35.71	91.70	92.56
Membranous nephropathy	35.00	93.04	92.56

(Averages over 100 runs.)

Table 3 Summary of Computer Analysis for all Individual Major Conditions

IF	Anorexia is True
	Urinary RBC Count >= 29546
	Urinary Total Protein <= 6.16
	Serum Creatinine >= 85.67
	Blood Urea <=30.8
	Serum Total Protein <= 73
	24 <= Serum Albumin <= 38
	Haemoglobin <= 13.4
	White Cell Count >= 4.0
	126 <= Platelets <= 495
	46 <= ESR<= 130
THEN	Diagnosis = Microscopic polyarteritis
IF	Upper Respiratory Infection is False
	Anorexia is True
	Weight Loss is True
	Bowel Disturbance is False
	Urinary RBC Count >= 27500
	84.80 <= Serum Creatinine <= 710.00
	Serum Albumin >= 23
	Haemoglobin <= 14.2
	Platelets >= 125
	Antinuclear Factor is Negative
THEN	Diagnosis = Microscopic polyarteritis

The lines between the keywords IF and THEN specify conditions. The Einstein system has inferred that if all of the conditions for either rule are satisfied then the patient has microscopic polyarteritis.

Figure 1. An example set of rules for diagnosing microscopic polyarteritis developed by the Einstein system.

1000 expert systems of the former type were developed from different subsets of the data-base. On average, these had a diagnostic accuracy of 51.05% when applied to cases that had not been used in their development. This can be compared with an average diagnostic accuracy of 48.31% obtained by 4 expert clinicians each of whom were asked to diagnose 30 randomly selected cases from the data-base on the basis of the information contained therein alone. It should also be interpreted in the context of a chance accuracy of 3% obtained by random selection between the 31 conditions represented in the study.

This result is especially good when it is considered that for each of more than half of the conditions present in the data-base there are 5 or fewer cases present and that one third of the conditions are represented by a single case each. Further, the data-base does not include many variables which are believed to be important, as detailed above.

References

[1] Funck-Bretano, J.L. (1989). Potential contribution of artificial intelligence to nephrology. *Contrib Nephrol* 70:22 - 29.

[2] Michalski, R. S. (1984). A theory and methodology of inductive learning, in Michalski R.S., Mitchell, T.M. and Carbonell, J.G (eds): *Machine Learning: An Artificial Intelligence Approach*. Berlin, Springer-Verlag, pp. 83-129.

[3] Michalski, R. S., Mozetic, I., Hong, J. & Lavrac, N. (1986). The multi-purpose incremental learning system AQ15 and its testing application to three medical domains. *AAAI-86: Proceedings of the Fifth National Conference on Artificial Intelligence*, pp. 1041 - 1045.

[4] Quinlan, J.R. (1987). Simplifying decision trees. *Int. J. Man Machine Studies* 27:22 1 - 234.

[5] Quinlan, J.R., Compton, P., Horn, K.A., et al. (1986). Inductive knowledge acquisition: a case study. *Technical report 86.4, NSWIT School of Computing Sciences*, Sydney.

[6] Tomura, S., Tsutani, K., Sakuma, A., et al (1985). Discriminant analysis in renal histological diagnosis of primary glomerular diseases. *Clin Nephrol.* 23:55-62.

[7] Urakabe, S., Orita, Y., Shirai, D., et al. (1975) Prognosis of chronic glomerulonephritis in adult patients estimated on the basis of the Markov process. *Clin Nephrol* 3:48-53.

[8] Webb, G. I. (1991a) Learning disjunctive characteristic descriptions by least generalization. *Technical Report 2/9 1, Department of Computing and Mathematics, Deakin University*.