Department of Computing Science and Mathematics
University of Stirling

Classifying Acute Abdominal Pain By Assuming Independence : A Study Using Two Models Constructed Directly From Data

Clifford S Thomas

Department of Computing Science and Mathematics, University of Stirling
Stirling FK9 4LA, Scotland

Telephone +44-786-467430, Facsimile +44-786-464551
Email cst@cs.stir.ac.uk

Technical Report CSM - 153

January 1999

Abstract

Medical Expert Systems have been under development for several years, MYCIN [1] being a typical example of a major project. Many of these recent systems have been constructed through expert opinions. In this paper we focus our attention on the semi-automated construction of a classifier directly from the database itself. The methodology employed utilises a data set specifically concerning the domain of Acute Abdominal Pain (AAP), courtesy of St John's Hospital, Livingston, Scotland. Our objective is : given a set of diseases the computer diagnostic system will use a certain procedure to classify patients in accordance with their symptoms.

The classifying procedure we propose to use centers on the assumption of independence. More specifically, a model described by naive Bayes (some times referred to as simple Bayes) and a new proposed model derived from Mutual Information Measure (MIM). Both models are constructed from a 2/3 randomly selected subset of the AAP data set and then validated by the remaining 1/3 unseen test data sample.

A total of 3705 patient record test samples were classified using the two models and compared to the results obtained by the Doctors at St John's Hospital. The MIM classifier identified correctly 70.04% of the unseen test samples with the naive Bayes classifier achieving 73.17%. In comparison to the Doctors' 70.77%, both models of independence performed with a similar level of expertise, naive Bayes being marginally higher.

Acknowledgements

I wish to thank the staff of St John's Hospital, Livingston for their permission to use the CADA (**C**omputer **A**ssisted **D**iagnosis & **A**udit) database and in particular Mrs Julie Mclaren for her help in understanding its contents. In addition I would like to thank Dr Yiqun Gu for her assistance with the G&T system and the implementation of the naive Bayes classifier used within this paper.

Contents

1 INTRODUCTION

EARLY work in Artificial Intelligence concerning the building of expert systems involved a tedious process of manual knowledge acquisition [2]. Through a procedure of assessment and elicitation experts were required to be interviewed in order to obtain prior estimates of relevant quantities. This presented knowledge engineers with a problem. Whilst on one hand, domain experts may be needed in order to circumscribe the learning component, for example, which variables might be used, what is being predicted from what, etc. On the other hand, domain experts can be poor at judging their own limitations and capabilities, and estimating probabilities [3]. However, to build a model and carry out subsequent evaluation to determine what's going on with the data, we often require prior knowledge and this may only be obtainable from domain experts.

This sort of dilemma led naturally to the development of systems whose elicitation processes suffered less constraints and uncertainty. Researchers such as Pearl [4] and Spiritus [5] proposed through their work some of the earliest algorithms for learning structure from data. Essentially, on the assumption that there is a sufficient volume of data, in order to accurately estimate various probabilities, a graphical structure with parameters can be reconstructed. This approach removes the knowledge acquisition bottleneck and reduces the dependency upon expert subjectiveness.

The theory of network identification from data forms a precursor to techniques for learning using representative samples. The earliest result in structure learning was the Chow and Liu algorithms for learning trees from data [6]. That is it learns a Bayesian network whose shape is a tree. Kutató[7] demonstrated further that more complex structure learning was possible from quite reasonable sample sizes.

Other early work on structure learning was often based on identification methods and the reader is directed towards [4,5,8,9] and to [34,35,36,37] for systems that extract knowledge from databases.

In this paper we present two models that represent methodologies of semi-automated expert system construction. Using a domain specific database describing AAP both models extract the structure and corresponding branch parameters, directly from the database. One model is founded upon a simplification of Bayes theorem, the other is a modification to Chow and Liu [6] and its tree construction algorithm. Both models will be investigated in respect of their ability to classify and compared to results obtained from domain experts on a sample of unseen test data.

2 HISTORY OF THE DATABASE - ST JOHN'S HOSPITAL

PRIOR to patient records being formalised into a standard format, during Accident and Emergency (A&E) admittance, the average success rate for diagnostic accuracy was around 55%. This resulted in 18% of all emergency abdominal operations being unnecessary with the patient being placed in unnecessarily high risk, discomfort and subsequent loss of working days.

The introduction of a standard format for collecting case history of patient symptoms and examination increased the casualty officers' diagnostic accuracy to 66%. Essentially a standard data collection form containing 33 data points covering 135 features was completed during a patients examination on arrival to the A&E department.

A natural progression from the manual completion of standard forms was the introduction of a computer assisted diagnostic system. Work previously carried out by Professor Tim de Dombal in the academic unit, Leeds [10,11] was continued by Mr AA Gunn at Bangour General Hospital. A computer assisted diagnostic program was introduced as routine practice into the A&E department. Casualty officers were required to enter the case history and examination, using the standard form as a source of input for all patients entering A&E. In return the computer would supply the probability of a disease (one from a possible nine).

With the examination procedures now structured it was possible to follow up cases to not only determine what actually happened to the patient but provide the casualty officers with feedback of their 'initial' diagnostic accuracy.

The introduction of the completed system together with an auditing procedure increased the diagnostic accuracy of the casualty officers to 76%.

The collective information gathered both during examination and subsequent audit administration formed what is considered the largest database of AAP in Europe and is abbreviated as CADA. The Computer Assisted Diagnostic and Audit database is essentially a collection of patient records for every abdominal pain attendance containing every symptom and investigation. Each completed file records the Doctors' 'initial' diagnosis, the computers suggested diagnosis (based upon a naive Bayes model) and the 'actual' diagnostic group a patient was determined as really belonging to, on their discharge from hospital.

2.1 The CADA Database - Description

The CADA database will be used to test the proposed methods. The total consists of 10,927 records of patients who were admitted to hospital suffering from acute abdominal pain. The actual contents of this database far exceed the requirements of this paper and mainly provide information necessary for Hospital Audits. The precise format relevant to this paper can be found in Appendix B. The CADA database consists of nine groups as defined by the experts concerning the domain of Acute Abdominal Pain (AAP). These are : Appendicitis (APP), Diverticulitis (DIV), Perforated Peptic Ulcer (PPU), Non Specific Abdominal Pain (NSAP), Cholecystitis (CHO), Intestinal Obstruction (INO), Pancreatitis (PAN), Renal Colic (RCO) and Dyspepsia (DYS). With the exception of one attribute namely 'AGE', which is strictly a continuous variable, all of the 32 attributes represent discrete variables, the Doctors themselves have provided the discrete parameter conversion. The group NSAP is not actually a diagnostic group but a catch all category into which the doctors assign the patients who do not fit into one of the other true eight diagnostic groups. PAN is a poorly characterised group

as many of the significant symptoms, such as a blood test for levels of alcohol, are not included within the 135 symptoms recorded during patient examination. For the proposes of this work, the database sample has been partitioned into an 'extraction/learn' and test block. The extraction/learn part being approximately 2/3 whilst the test, the remaining 1/3 of the entire sample size. The data is randomly split into the two partitions. One comprising of 7222 training samples (2/3) and the remaining 3705 samples representing unseen test data (1/3 external). Table 1 shows the actual data distributions.

Disease Group	Training Data	Test (External) Data	Totals
APP	770	385	1,155
DIV	182	92	274
PPU	112	56	168
NSAP	3,344	1,764	5,110
CHO	614	308	922
INO	456	228	684
PAN	144	72	216
RCO	573	287	860
DYS	1,025	513	1,538
Totals	7,222	3,705	10,927

Table 1. Sample Distributions.

With reference to Appendix B, overall there are 33 symptoms and 9 diagnostic groups or classes. Each symptom has associated with it a set of parameters, for example symptom number 21 MOOD has parameters : normal (21/0), distressed (21/1), and anxious (21/2). In total there are 135 parameters describing the 33 symptoms. In a similar manner to the symptoms, diagnostic groups have 2 parameters namely, Absent and Present.

On examination of the database, records were found to have multiple parameter values stored in respect of some of the symptoms and in other cases none of the symptoms parameters were recorded. Rather than ignore these anomalies, two additional parameter values have been attached to each of the symptoms. They are : missing (99) and composite (88). For example MOOD number 21 will now be described by parameters normal (21/0), distressed (21/1), anxious (21/2), composite (88) and missing (99). This increases the input stream from 135 to a total of 201.

It is possible to gather groups of symptom parameters that have been assigned to composite (88) attributes, and represent them as a 'single' combinatorial parameter in respect of each symptom. Assuming that the data representing these combinations of parameters is not too sparse. In deed the G&T [12] system actually considers symptom parameters as a series of combinations, however, in this system c^2 thresholding is employed to determine irrelevance and overcome the problems of scarcity. In this paper, we will be assuming models of independence. Symptom parameters grouped in this way (considered to be significant by c^2 tests) imply dependency and so conditional independence is no longer valid. In addition a

symptom consisting of n parameters, would mean that a symptom parameter list would expand to 2^n - 1 (missing (99) items not being included), resulting in a rise of complexity. With this in mind we shall avoid using this form of representation and assign (88) to all composites.

3 MODELS OF INDEPENDENCE

WHEN a diagnosis is performed**,** and it is necessary to take into account more than one symptom, the situation becomes complicated. However, a simplification is possible if it can be assumed that certain symptoms are independent of each other. Previous work that assumed a model of independence proved both to perform well, sometimes outperforming conventional dependency models, and to reduce the computational complexity in dealing with dependencies. Professor Tim de Dombal's system [10,11], which assumed independence in naive Bayes, successfully applied the theorem to the diagnosis of Acute Abdominal Pain diseases. Another was PROSPECTOR [13] which was designed to aid geologists in evaluating mineral sites for potential ore deposits. The reader is directed to Michie [14,15] for a detailed study comparing the naive Bayes classifier to other learning algorithms.

The development of both these systems was based on a set of controversial assumptions.

- The symptoms must be independent. That is, the appearance of one symptom cannot make a second symptom more likely.
- The disease set must be complete. This assumes that a patient will have a diagnostic outcome that must be one of the diseases represented by the system.
- The diseases must be mutually exclusive. It assumes a patient can have one and only one disease.

A set of *n* hypothesis is said to be mutually exclusive with respect to *I* if

$$P(H_i, H_j \mid I) = 0 \;\; for \;\; i \neq j \;\; where \; I \; is \; background \; Information \left(content \right).$$

Two models of independence are presented in this paper and are both evaluated using a database describing records of patients suffering from AAP. The first is a simplification of Bayes theorem, the second a new approach utilising Mutual Information Measure. Both build a classifier from 'extraction/learning' data and subsequently use unseen test samples to assess their individual performances.

4 BAYES THEOREM

THE Rev. Thomas Bayes was an 18th century mathematician who derived a special case of this theorem [16]. The theorem was generalised by Laplace [17], and represents the basic starting point for inference problems using probability theory as logic [18].

Machine learning researchers are often interested in determining the best hypothesis from some space *H*, given the observed training data *D*. One way to specify this best hypothesis is to demand the most probable hypothesis, given the data *D* together with any initial knowledge about the prior probabilities of the various hypothesis in *H*. Bayes theorem provides a direct method for calculating such probabilities. More precisely, Bayes theorem provides a way to

calculate the probability of a hypothesis based on its prior probability, the probabilities of observing various data given the hypothesis, and the observed data itself.

The Bayesian approach to classifying a new instance is to assign the most probable target value V_{map} (maximum a posteriori), given the attribute values $(a_1, a_2, \ldots, a_n)$ that describe the instance.

$$V_{map} = \arg\max_{V_j \in V} \frac{P(a_1, a_2, \ldots a_n \mid V_j) P(V_j)}{P(a_1, a_2, \ldots, a_n)}$$

Using Bayes theorem we can rewrite this as

$$V_{map} = \arg\max_{V_j \in V} \frac{P(a_1, a_2, \ldots a_n \mid V_j) P(V_j)}{} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Using the training data, we can estimate $P(v_j)$ simply by converting the frequency with which each target value v_j occurs in the training data. Estimating the different $P(a_1, a_2, \ldots, a_n \mid v_j)$ terms in the same way is not however feasible without a great amount of training data. Since the number of these terms is equal to the number of possible instances times the number of possible target values, then we would need to see every instance in the instance space many times in order to obtain reliable estimates.

4.1 Naive Bayes - Model I

The naive Bayes (NB) classifier is based on the simplifying assumption that the attribute values are conditionally independent given the target value. That is, the assumption is that given the target value of the instance, the probability of observing the conjunction $a_1, a_2, \ldots, a_n$ is just the product of the probabilities for the individual attributes :

$$P(a_1, a_2, \ldots, a_n \mid v_j) P(v_j) = \prod_i P(a_i \mid v_j)$$

If we substitute this into (1) we get :

$$V_{NB} = \arg\max_{V_j \in V} P(v_j) \prod_i P(a_i \mid v_j) \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

where $P(a_i \mid v_j)$ terms can be estimated from the training data and is just the number of distinct attribute values times the number of distinct target values. Thus $P(v_j)$ and $P(a_i \mid v_j)$ terms are estimated using their frequencies over the training data. Essentially, the set of estimates corresponds to the 'learnt' hypothesis and is used to classify each new instance by applying the rule of equation (2) above.

To classify AAP we require a model that can deal with observed patterns of symptoms, and through the model determine the most probable disease or posterior probability. If we let *P(H)* denote the initial probability that hypothesis *H* holds, before we have observed the training data. (prior probability). *P(E)* denote the prior probability that training data *E* (evidence pattern) will be observed and *P(E|H)* denote the probability of observing data *E* given some world in which hypothesis *H* holds then Bayes theorem provides a way to calculate the posterior probability *P(H|E).*

That is

$$P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)} \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

now if *E* is described by factors such as symptom parameters giving $(E_1, E_2, \ldots\ldots E_n)$ then

$$P(E) = P(E_1 \mid H) \cdot P(E_2 \mid E_1, H) \cdot P(E_3 \mid E_1, E_2, H) \ldots\ldots\ldots\ldots\ldots\ldots (4)$$

and

$$P(E) = P(E_1) \cdot P(E_2 \mid E_1) \cdot P(E_3 \mid E_1, E_2) \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

so

$$P(H \mid E) = P(H) \cdot \frac{P(E_1 \mid H)}{P(E_1)} \cdot \frac{P(E_2 \mid E_1, H)}{P(E_2 \mid E_1)} \ldots\ldots\ldots\ldots\ldots\ldots (6)$$

In the case of the naive Bayes classifier, we assume that the instance attribute E_1 is conditionally independent of the instance attribute E_2 given the target value *H*.

Under this assumption equation (6) becomes simplified to :

$$P(H \mid E) = P(H) \cdot \frac{P(E_1 \mid H)}{P(E_1)} \cdot \frac{P(E_2 \mid H)}{P(E_2)} \cdot \frac{P(E_3 \mid H)}{P(E_3)} \ldots\ldots\ldots\ldots\ldots\ldots (7)$$

4.2 Application of Naive Bayes to CADA Database.

The ‘extraction/learning’ partition was used to construct the naive Bayes model and the unseen test data used to validate it. The model in this particular case being the calculation of the marginal and conditional probabilities using the information contained in the database, Equation (7) now represents an equivalent discriminator accepting the test data as input and producing nine probabilities as its output. In this case the highest probability, corresponding to a diagnostic group, is taken to be the resulting classification in respect of a particular test sample presented to the model. Table 2 displays the resulting classification for naive Bayes, with the off-diagonals indicating the miss-classifications.

Disease Group	Classification Output Results									
	APP	**DIV**	**PPU**	**NSAP**	**CHO**	**INO**	**PAN**	**RCO**	**DYS**	**Totals**
Input Vector										
Sample for APP	283	3	5	82	0	1	2	1	8	385
Sample for DIV	0	48	5	19	1	11	1	3	4	92
Sample for PPU	0	0	41	0	5	2	7	0	1	56
Sample for NSAP	130	44	8	1,348	23	70	4	69	68	1,764
Sample for CHO	1	0	7	12	192	22	15	4	55	308
Sample for INO	9	16	6	25	8	140	12	4	8	228
Sample for PAN	1	2	8	3	7	5	23	2	21	72
Sample for RCO	7	7	1	29	8	3	2	228	2	287
Sample for DYS	2	9	7	30	20	16	20	1	408	513
Totals	433	129	88	1,548	264	270	86	312	575	3,705

Table 2. The Test Data Set Results 73.17%

4.3 Discussion of Results.

The comparative model of independence, defined here by naive Bayes, can be seen to perform extremely well. Its performance as a classifier shows it achieving 73.17% for the unseen test sample. Clearly an indication that as model of 'independence' it can be very efficient, particularly as many an expert would argue that a model of dependence should be a better classifier, especially for medical data applications.

5 A MUTUAL INFORMATION MEASURE MODEL

SHANNON'S information theory [19] has spawned many artificial intelligence applications particularly those of the machine learning community. **ID3** [20,21] designed for dealing with problems where there are many attributes and the training set contains many objects, and **C4.5** [22] its direct descendant, are prominent examples of information measure utilisation. **CART** [23] is another example of employing in its initial development the information measure. This system was constructed to solve problems met by designers of some inductive learning systems, with the objective of deriving efficient node splitting criteria. In fact the use of mutual information has been found in the selection of features in supervised neural network learning [24] and for discovering dependencies in DNA sequences [25].

In this paper a new model of independence is presented which will use the mutual information measure to extract the 'structure' of a 9 class group directly from the CADA database. This is essentially an extension to work carried out on a subset of this database [26].

5.1 Learning Structure from Data

Network construction directly from empirical observations has been proposed by Pearl [27]. The basis of construction lies in first extracting the undirected tree using the Chow/Liu algorithm [6] and then determining the corresponding branch parameters.

Chow/Liu proposed that in order to determine the best tree dependent distribution that approximates an estimated or measured distribution the Kullback-Leiber cross entropy measure [28,29] could be chosen as a distance criterion between distributions. For the two distributions P and $P^{'}$:

$$D(P,P^{'}) = \sum_{x} P(x)\log\frac{P(x)}{P^{'}(x)} \quad \ldots\ldots\ldots\ldots\ldots\ldots (8)$$

That is, amongst all the spanning trees that one can draw on n variables, each yielding a product P^t, the closest P^t to the measured or estimated P can be determined. The measure is non-negative and attains a value 0 if and only if $P^{'}$ coincides with P. The distance measure of (8) can be minimised by projecting P on any maximum weight spanning tree (MWST), where the weight on the branch (X_i , X_j) is defined by the mutual information measure :

$$I(X_i, X_j) = \sum_{X_i, X_j} P(X_i, X_j)\log\frac{P(X_i, X_j)}{P^{'}(X_i, X_j)} \geq 0 \quad \ldots\ldots\ldots\ldots\ldots\ldots (9)$$

Similarly the mutual information between the two variables X_i and X_j can be represented by :

$$I(X_i, X_j) = \sum_{X_i, X_j} P(X_i, X_j)\log\frac{P(X_i, X_j)}{P(X_i)P(X_j)} \quad \ldots\ldots\ldots\ldots\ldots\ldots (10)$$

Equation (10) will be the representation employed for the purpose of generating the model.

5.2 Chow/Liu and Structure Extraction

With the assumption of symptom independence, discussed earlier, the mutual information measure was calculated by applying (10) to all variable couples for a training data set of 7222 items. Initially both symptom and diagnostic groups are considered equally, resulting in 9 + 33 variables. (The symptom parameters are summed with respect to each symptom block). For each diagnostic group the corresponding branch mutual information measures were collected generating a single diagnostic tree structure as shown in figure 1.

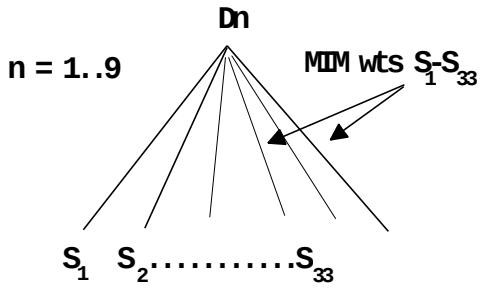

Figure 1 – Individual Structure

Thus nine trees will be constructed, one for each of the diagnostic groups.
In each case only the non-negative values of information measure are used with negative values being discarded as invalid.

Figure 2 represents the constructed mutual information measure model.

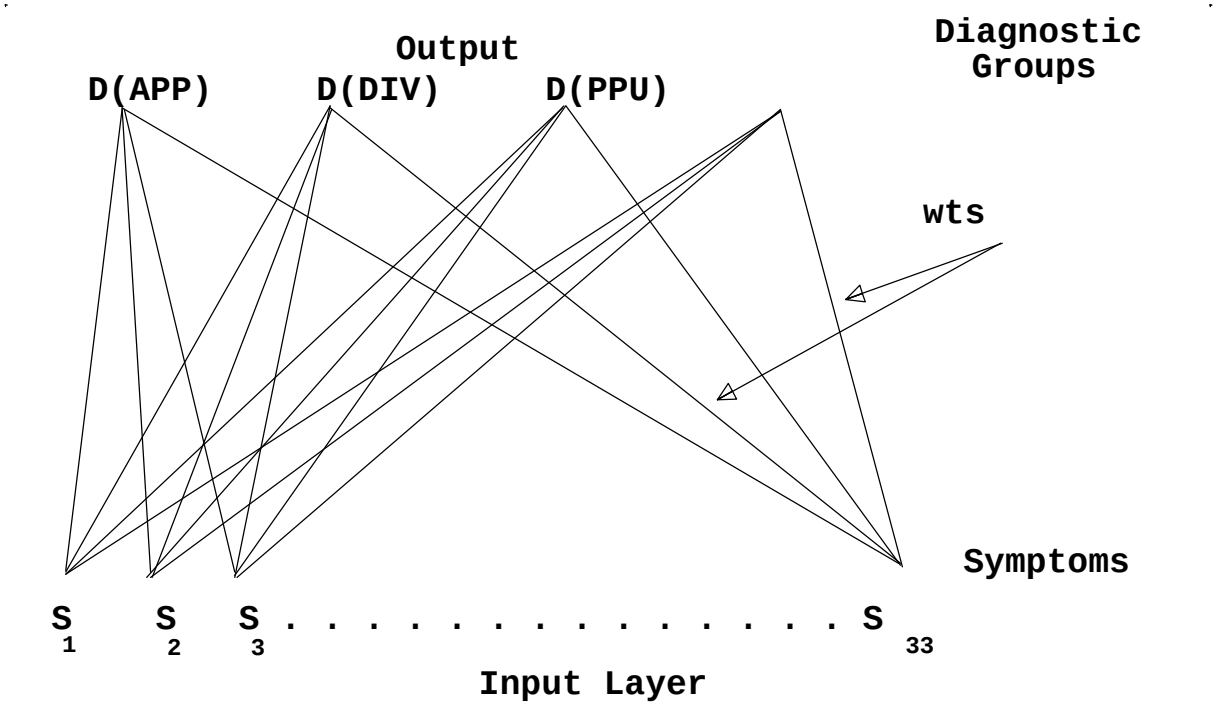

Figure 2 – MIM Model Nine Group Structure

5.3 The MIM Model as a Classifier - Model II

In order to define a classifier, We propose to consider the MIM as a form of weight. With respect to each branch connecting a symptom to a particular diagnostic group the calculated value of non-negative information measure will represent the strength or weight for that particular branch. Similarly for the other model branches. Since a model of independence is assumed there will be no branch connections between symptoms or between diagnostic groups.

The tree(s) constructed optimise to obtain a structure such that the tree sum of the mutual information is maximised. That is, for n significant pairs :

$\sum_{i=1}^{n} \hat{I}(X_i, X_j)$ *where* $\hat{I}(X_i, X_j)$ is the mutual information measure ……(11)

In terms of classification, the algorithm maximises the likelihood function representing a maximum likelihood estimator (MLE) for the dependence tree. To result in a classification, the input data must match one of the trees describing the diagnostic groups. That is, define the maximum likelihood estimation for a particular dependence tree.

Thus multiplying each corresponding branch mutual information measure 'weight' by its associated input value, results in the total depicted by (11). Subsequently an equivalent linear discriminate function can be utilised as a means by which a decision can be made. This represents a similar concept to that employed by Gallant [30,31] in his connectionist Expert System. That is :

$$S_I = \frac{1}{I_{\max}} \sum_{i=1}^{n} \hat{I}(X_i, X_{j(i)}) \ldots\ldots\ldots\ldots\ldots\ldots (12)$$

where $I_{\max} = \sum_{Tree} I(X_i, X_{j(i)})$ for each particular diagnostic tree, and $I(X_i, X_{j(i)})$ is calculated from (10).

Since the diagnostic groups are not equal in sample size, the value Imax acts as a group 'balance' unifier.

Thus the resulting total depicted by (11) can achieved by multiplying (12) by the associated input value :

$$S_I = \frac{1}{I_{\max}} \sum_{i=1}^{n} \hat{I}(X_i, X_{j(i)}) \times input \ldots\ldots\ldots\ldots\ldots\ldots (13)$$

where $S_I > 0$ defines a TRUE outcome.

Clearly (13) has associated within its definition a strength or ordering. By taking the highest positive value of *'S'* amongst the four diagnostic groups, the most likely outcome can be determined.

A direct comparison of (13) with Gallants' linear discriminant function suggests that a substitution of the mutual information measure branch weights could be made for those Gallant determined through use of the pocket algorithm [30].

5.4 Application of MIM Classifier to CADA Database.

Applying the Chow/Liu algorithm [6] and assuming a model of independence, the extraction data was successfully transformed into a single layer tree structure with MIM weights attached to each individual branch. In fact nine trees were generated, each one representing one of the nine diagnostic groups. Using this structure the unseen test data was inputted into the model with the resulting classification achieved as shown in table 3, whilst table 4 the results from the 'extraction' partition data.

The level of success for the entire test data set was 69.96% and the 'extraction' data set 66.2% The shadowed diagonal represents the classification results that were correctly identified, with the level of miss-classification indicated by the off-diagonal scores.

Classification Output Results

Disease Group	APP	DIV	PPU	NSAP	CHO	INO	PAN	RCO	DYS	Totals
Input Vector										
Sample for APP	286	0	9	79	0	0	4	2	5	385
Sample for DIV	6	23	7	38	1	11	4	0	2	92
Sample for PPU	2	0	42	2	4	1	3	0	2	56
Sample for NSAP	125	22	14	1,418	18	63	11	50	43	1,764
Sample for CHO	6	1	10	40	148	15	20	3	65	308
Sample for INO	14	2	11	33	4	128	17	7	12	228
Sample for PAN	3	1	12	5	3	2	18	2	26	72
Sample for RCO	10	5	1	60	8	4	4	195	0	287
Sample for DYS	6	5	7	102	14	13	23	9	334	513
Totals	458	59	113	1,777	200	237	104	268	489	3,705

Table 3. The Test Data Set Results 69.96%.

Classification Output Results										
Disease Group	**APP**	**DIV**	**PPU**	**NSAP**	**CHO**	**INO**	**PAN**	**RCO**	**DYS**	**Totals**
Input Vector										
Sample for APP	590	0	22	141	2	3	2	3	7	770
Sample for DIV	8	42	15	74	2	33	4	2	2	182
Sample for PPU	4	2	75	11	4	1	5	1	9	112
Sample for NSAP	511	44	30	2,416	36	104	34	79	90	3,344
Sample for CHO	12	1	25	63	316	22	25	11	138	613
Sample for INO	19	19	16	90	3	269	18	9	13	456
Sample for PAN	5	1	18	11	15	8	37	4	45	144
Sample for RCO	24	6	4	126	7	8	8	389	1	573
Sample for DYS	11	4	15	194	45	19	76	18	643	1,025
Totals	1,184	119	220	3,126	430	467	209	516	848	7219*

Table 4. The 'Extraction' Data Set Results 66.2%.

* Note: Three patient records were found to have several missing symptoms which resulted in NSAP and CHO failing to classify them into any of the nine groups, thus being lost.

5.5 Discussion of the Results

With the nine diagnostic groups defining the full CADA classifier, the poorly characterised groups clearly have suffered in respect of the other groups. The overall classification for the unseen test data was 69.96% which is lower than that achieved when only four of the groups was first investigated [26]. PAN and DIV being particularly low.

In the case of the extraction partition data set, the overall classification was 66.2% which is less than the result obtained for the test sample. This perhaps suggests there is room for some improvement through optimisation and that the influence of NSAP, that is its generalised structure, is strong enough to steal many of the other group individual 'true' test samples.

6 REDUCING THE MODEL COMPLEXITY

STUDIES and applications utilising contingency tables have long been a part of statistical analysis. Kullback [28] in conjunction with the concepts of communication theory proposed that the significance c^2 [32, 33] could be approximated by the independence component $\hat{I}(H_1 : H_2)$ multiplied by 2. H_2 is the null hypothesis and H_1 the alternative. That is :

$$H_2 : P_{ij} = P_{i.}P_{.j} \text{ and } H_1 \text{ the alternative } H_1 : P_{ij} \neq P_{i.}P_{.j}$$

If we consider in particular the two - way table then we have

$$2\hat{I}(H_1 : H_2) \approx \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{\left(x_{i,j} - \frac{x_{i.}x_{.j}}{N}\right)}{\frac{x_{i.}x_{.j}}{N}} \sim c^2 \ldots\ldots\ldots\ldots\ldots\ldots (14)$$

for *(r-1)(c-1)* degrees of freedom.

Where for *N* independent observations x_{ij} is the frequency of occurrence in the ith row and jth column and :

$$x_{i.} = \sum_{j=1}^{c} x_{ij},\ x_{.j} = \sum_{i=1}^{r} x_{ij} \text{ and } N = \sum_{i=1}^{r}\sum_{j=1}^{c} x_{ij}$$

Denoting probability by *P* with corresponding subscripts that is :

$$P_{ij} = \frac{x_{ij}}{N},\ P_{i.} = \frac{x_{i.}}{N}, \text{ and } P_{.j} = \frac{x_{.j}}{N}$$

Then equation (14) can be written as :

$$2\hat{I}(H_1 : H_2) = \sum_{i=1}^{r}\sum_{j=1}^{c} NP_{ij} \log \frac{P_{ij}}{P_{i.} \times P_{.j}} \ldots\ldots\ldots\ldots\ldots\ldots (15)$$

which is :

$$2\hat{I}(H_1 : H_2) = 2\sum_{i=1}^{r}\sum_{j=1}^{c} x_{ij} \log \frac{Nx_{ij}}{x_{i.}x_{.j}} \approx c^2 \ldots\ldots\ldots\ldots\ldots\ldots (16)$$

using (16), equation (10) can be rewritten as :

$$I(X_i, X_j) = \sum_{x_i, x_j} P(X_i, X_j) \log \frac{P(X_i, X_j)}{P(X_i) \times P(X_j)} \text{ and so :}$$

$$2 \times I \times N \approx c^2 \ldots\ldots\ldots\ldots\ldots\ldots (17) \text{ for df } \textit{(r-1)(c-1)}$$

By applying equation (17) to the mutual information measure model, the input variables that were found to be independent, in respect of each diagnostic group could be removed. That is, the mutual information *I(X_i , X_j)* between attribute and class can be used to judge if the attribute could, of itself, contribute usefully to a classification scheme thereby offering a measure of irrelevance. Non-significant or irrelevant branches were identified using five different c_c^2 levels of confidence (c = .9, .95, .955, .99, .995). Each of these modified models were again tested with the unseen test data samples. The resulting classifiers for both thresholding and non-thresholding did not appear to be significantly different. However, the models after thresholding did have a reduction of the number of input variables, although

only visible at the individual structure levels, and thus an overall reduction in model complexity.

The tables in Appendix A illustrate the levels of misclassification (off - diagonals) for each of the threshold in respect of the 3705 test samples.

6.1 Discussion of Results

In general the application of kullback thresholding to the MIM model did not improve the overall classification results dramatically. The best or optimal obtained was 70.04% for $c^2_{.99}$. However, what kullback did achieve was a reduction in the model complexity and thus a reduction in the redundancy with out loss of performance.

It was clearly indicated that for all levels of confidence applied to the 'extraction' partition, symptoms 2, 3 ,4, 26 and 32 were found to be significant for all of the nine diagnostic groups. These particular symptoms will have very high values of mutual information measure in respect of their coupling with the nine groups, and thus will have very significant branch weights.

7 THE DOCTORS 'INITIAL' CLASSIFICATION - THE EXPERTS

WHEN a patient is examined by a Doctor, a record of the symptom parameters together with the Doctors' initial diagnosis is recorded and stored in the Hospital CADA database. This can be considered as the Experts classification results and can be used as a comparitor for the MIM and naive Bayes resulting classifications. Table 5 displays the level of success for the test partition sample data.

In the doctors case, the resulting classification output groups are not restricted to the nine AAP ones selected to do this research (i.e. the CADA database). The data set represents a sample with known resulting diagnostic group assignments as recorded on a patients discharge from hospital. That is, the actual disease attributed to a particular set of symptoms for a particular patients is already known as a fact.

When a doctor however, examines a patient, no restrictions are imposed to force a decision to fall into these nine categories during the 'initial' diagnostic assignment. Those that do fall outside the nine groups are considered still considered as miss-classifications, and the table of results thus shows a total sample size less than the actual test size of 3705.

	Classification Output Result									
Disease Group	**APP**	**DIV**	**PPU**	**NSAP**	**CHO**	**INO**	**PAN**	**RCO**	**DYS**	**Totals**
Input Vector										
Sample for APP	308	3	3	42	3	2	2	2	2	367
Sample for DIV	1	47	6	11	3	13	0	1	1	83
Sample for PPU	0	0	41	1	5	2	1	0	4	54
Sample for NSAP	211	23	5	1,169	17	44	6	42	31	1,548
Sample for CHO	3	2	8	10	212	10	14	3	34	296
Sample for INO	7	3	2	14	2	174	3	1	7	213
Sample for PAN	0	2	6	5	4	4	45	1	3	70
Sample for RCO	5	1	1	7	5	3	1	240	2	265
Sample for DYS	3	4	10	44	27	7	19	0	386	500
Totals	538	85	82	1,303	278	91	259	290	470	3396*

Table 5. The Test Data Set Results 70.77%.

*Note: 309 items have been miss-classified into groups outside the nine used.

7.1 Discussion of Results

For the randomly selected test partition used within this paper table 5 displays the level of classification that the Doctors achieved on their ***initial*** estimate of the patients disease group. At 70.77% this is a representative level compared to the MIM, but less than naive Bayes. What is particularly interesting is the PAN and DIV individual levels of classification success. DIV is essentially as high as that in naive Bayes but PAN is considerably better than either the MIM or naive Bayes classifiers. Clearly since the database does not define PAN fully, the Doctors must be using heuristics for these diagnosis, as they would not have had the opportunity to do blood tests prior to their initial input to the CADA database record.

It should also be noted that the Doctors are not restricted to the nine groups that have been selected for Acute Abdominal Pain, that is this database subset. Female patients have other categories for which many patients can be assigned. Therefore in the case of the Doctors classification totals of 3396 it indicates that 309 patients may have been classified outside the nine groups and thus falling short of the 3705 test partition.

8 COMPARISON OF THE MODELS/EXPERTS CLASSIFICATIONS

TABLE 6 displays the classification results for all tests carried out within this paper. The clear winner is naive Bayes, with the remainder close behind averaging 70%.
In table 7 the three classifiers, one being the human expert, are compared in respect of the individual nine groups.

As already discussed Pancreatitis (PAN) is generally diagnosed when patients with gallstones, or those who are suspected alcoholics, are given blood tests. Since these symptoms and tests are not recorded in the database the resulting poor success rate is easily explained. However, this is only true for the two statistical models as the experts were able to apply heuristics and perform considerably better. The statistical classifiers failing due to the lack of data.

In table 7 the group classification for NSAP is particularly high for the MIM and closely followed by naive Bayes. In the case of the experts this is not so high. It would appear that the Doctors are reluctant to assign NSAP to patients but would rather suggest one of the other eight groups. Statistically this is not the case and the two models which depend upon data are thus performing better here.

Doctor	**Test Data**	**Extract. Data**	**Test** $c^2_{.95}$	**Test** $c^2_{.9}$	**Test** $c^2_{.97}$	**Test** $c^2_{.99}$	**Test** $c^2_{.995}$	**naive Bayes**
70.77%	69.96%	66.20%	69.9%	69.7%	70.1%	70.04%	70.1%	73.17%

Table 6. The Overall Test Data Set Results.

Group	**MIM (%)** $c^2_{.99}$	**Doctor (%)**	**naive Bayes (%)**	**Sample Size**
APP	74	80	73.5	385
DIV	27.2	51.1	52.2	92
PPU	75	73.2	73.2	56
NSAP	80.8	66.3	76.4	1,764
CHO	47.7	68.8	62.3	308
INO	54.8	76.3	61.4	228
PAN	25	62.5	31.9	72
RCO	68.3	83.6	79.4	287
DYS	64.5	75.2	79.5	513
Overall (%)	**70.04**	**70.77**	**73.17**	**3,705**

Table 7. The Test Data Set Results - Three classifiers.

The MIM has not performed quite as well as naive Bayes but does match that of the experts. In general all models are in the 70% area for the unseen test data. Where the doctors clearly out perform both statistical systems in the group PAN, however, it would appear that the experts have not relied on the data alone but used heuristics which cannot be considered a comparable factor in this case.

9 FURTHER WORK

THE group NSAP not only contains the largest distribution of data samples but is generally the group into which many of the test samples are assigned. This is true for both correctly identified and many miss-classified group samples.

NSAP is a generalised group which is used to describe many AAP characteristics, such as stress, constipation and other non-specific areas. It is essentially a catch all category and used when a patients symptoms and examination results are such that the doctor cannot assign one of the eight true groups normally attributed to AAP.

The current MIM classifier treats NSAP as a real group, however, it is possible to model the classifier in a similar manor to the methodology adopted by the experts. That is, we only consider eight groups as the output diagnostic groups and attempt to identify the NSAP test samples from within each of these eight groups. This can be approached by assuming that the Chow & Liu extraction, together with Kullback thresholding, extracts the structure of each diagnostic group with relatively distinct features. Thus NSAP samples should be easily identified as not belonging to any of the true eight structures. Research into exactly how to recognise the NSAP samples within the eight diagnostic groups will be the principal focus for further work.

10 CONCLUDING REMARKS

THIS paper described two classifiers that were modelled on the assumption of independence. Both classifiers were generated directly from the domain database describing Acute Abdominal Pain, and subsequently evaluated in order to determine their diagnostic capability. The results were compared with those obtained from the experts themselves and found to be of a similar order of success. The MIM proposed classifier like the naive Bayes classifier have each illustrated that the assumption of independence can be effectively used to represent a classifier for the domain AAP. In addition the proposed MIM classifier has demonstrated a successful addition to the family of classifiers, and offers a promising mechanism for semi-automated expert system construction.

References

[1] **Buchannan, B.G Shortliffe, E.H**: *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, (1985), Addison-Wesley Publishing Company.

[2] **Hayes-Roth, F Waterman, D.A Lenat, D eds**: *Building Expert Systems*, (1983), Addison-Wesley Publishing Company.

[3] **Kahneman, D Slovic, P Tversky, A**: *Judgement under Uncertainty:Heuristics and Biases*, (1982), Cambridge University Press.

[4] **Pearl, J Verma, T.S**: *A Theory of Inferred Causation: Principles of Knowledge Representation and Reasoning*, (1991), Morgan Kaufmann, pp 441 - 452.

[5] **Spirtes, P Glymour, C**: *An Algorithm for Fast Recovery of Sparse Causal Graphics*, (1991), Social Science computing Reviews, Vol 9, No 1, pp. 62 - 72.

[6] **Chow, C.K Liu, C.N**: *Approximating Discrete Probability Distributions with Dependence Trees*, Uncertainty in Artificial Intelligence, (1989), No 3, pp. 129 - 147

[7] **Herskovits, E.H Cooper, G.F**: *Kutató : An Entropy-Driven System for Construction of Probabilistic Expert Systems from Databases*, (1990), Proc. sixth Conf. Uncertainty in Artificial Intelligence, pp. 54 - 62.

[8] **Fung, R.M Crawford, S.L**: *A System for Induction of Probabilistic Models*, (1990), Eighth Nat'l Conf. Artificial Intelligence, pp. 762 - 779.

[9] **Srinivas, S Russell, S Agogino, A**: *Automated Construction of Sparse Bayesian Networks*, (1991), Artificial Intelligence Frontiers in Statistics, pp. 182 - 201.

[10] **de Dombal, F.T Leaper, D.J Horrocks, J.C Staniland, J.R McCann, A.P,** *Computer-aided Diagnosis of acute abdominal pain*, (1972), British Medical Journal Vol 2 pp. 9 - 13

[11] **de Dombal, F.T Leaper, D.J Horrocks, J.C Staniland, J.R McCann, A.P,** *Human and Computer-Aided diagnosis of acute abdominal pain: Further report with the emphasis on performance of clinicians* (1974), British Medical Journal Vol 1 pp. 376 - 380

[12] **Gammerman, A.J Thatcher, A.R**: *Bayesian Inference in an Expert System without Assuming Independence*, (1988), Technical Report 88/11, Heriot-Watt University.

[13] **Gashnig, J** : *PROSPECTOR: an expert system for mineral exploration*. in Michie, D ed. Introductory Readings in Expert Systems, (1982), Gordon and Breach Science Publishers, NY.

[14] **Michie, D Spiegelhalter**, **D.J Taylor, C.C eds**: *Machine Learning, Neural and Statistical Classification*, Ellis Horwood Series in Artificial Intelligence, (1994).

[15] **King, R.D Henery, R Feng, C Sutherland, A**: *A Comparative Study of Classification Algorithms: Statistical, Machine Learning and Neural Network,* Machine Intelligence No 13, (1994), pp. 311 - 359

[16] **Bayes, Rev T,**: *An Essay Toward Solving a Problem in the Doctrine of Chances*, Philos. Trans. R. Soc. London 53, (1763), pp. 370 - 418; reprinted in Biometrika 45, (1958), pp. 293 - 315, and Two Papers by Bayes, (1963), New York, Hafner.

[17] **Laplace, P.S**: *A Philosophical Essay on Probabilities*, unabridged and unaltered reprint of Truscott and Emory translation, (1951), Dover Publications, Inc., New York.

[18] **Bretthorst, G.L**: *An Introduction to Model Selection Using Probability Theory as Logic*, in Maximum Entropy and Bayesian Methods, (1994), Kluwer Academic Publishers, Dordrecht the Netherlands.

[19] **Shannon, C.E Weaver, W**: *The Mathematical Theory of Communication*, (1949), Urbana, IL: University of Illinois Press

[20] **Quinlan, J.R**: *Learning efficient classification procedures and their application to chess and games*, (1983), Machine Learning, eds. Michalski, R.S Carbonell, J.G and Mitchell, T.M, Tioga, Palo Alto, California.

[21] **Quinlan, J.R**: *Induction of Decision Trees*, Machine Learning , 1, (1986) pp. 81 - 106

[22] **Quinlan, J.R**: *C4.5- Programs for Machine Learning*, (1993), Morgan Kaufmann Publishers, San Mateo, California

[23] **Breiman, L Friedman, J.H Olshen, R.A Stone, C.J**: *Classification and Regression Trees*, (1984), Belmont, California: Wadsworth Int. Group

[24] **Battiti, R**: *Using Mutual Information for Selecting Features in Supervised Neural Net Learning*, IEEE Trans. on Neural Networks, (1994), vol. 5, No 4 pp 537 - 550

[25] **Milosavljevic, A**: Discovering dependencies via Algorithmic Mutual Information: A case study in DNA sequence comparisons, (1995), Machine Learning, Vol 21, No 1, pp. 35 - 50.

[26] **Thomas, C.S**, *A Mutual Information Measure Classifier* (1996), Technical Report CSM-137, Department of Computing Science & Mathematics, University of Stirling.

[27] **Pearl, J**: *Probabilistic Reasoning in Intelligent Systems*, (1988), San Mateo, CA: Kaufmann.

[28] **Kullback, S Leiber, R.A**: *On Information and Sufficiency*, Ann. Math. Statistics, (1951), vol. 22 pp 79 - 86

[29] **Kullback, S**: *Information Theory and Statistics*, (1968), New York, Dover Publishers

[30] **Gallant, S.I** : *Connectionist Expert System*, Communications of the ACM, 31 (1988), pp. 152 - 169

[31] **Gallant, S I**: *Neural Network Learning and Expert Systems*, The MIT Press, Cambridge, Massachusetts, (1994).

[32] **Moroney, M.J**: *Facts from Figures*, (1974), chapter 15, Pelican, pp. 246 - 270

[33] **Dixon, W.J Massey Jr, F.J**: *Introduction to Statistical Analysis*, Chapter 13, pp. 238 - 241

[34] **Schmitz, J Armstrong, G Little, J.D.C**: *Coverstory-automated news finding in marketing*, in DSS Transactions, Institute of Management Sciences, (1990), Providence, RI

[35] **Hoschka, P Klosgen, W**: *A support System for Interpreting Statistical Data*, in Knowledge Discovery in Databases, (1991), Cambridge, MA: AAAI/MIT, pp. 325 - 345

[36] **Piatetsky-Shapiro, G.F Matheus, C.J**: Knowledge Discovery Workbench: *An exploratory environment for discovery in business databases*, in Workshop Notes from the 9th National Conference on Artificial Intelligence: Knowledge Discovery in Databases, (1991), Anaheim,CA pp. 11 - 24

[37] **Matheus, C.J Chan, P.K Piatetsky-Shapiro G**: *Systems for Knowledge Discovery in Databases*, IEEE Trans. on Knowledge and Data Engineering, (1993), vol. 5, No 6 pp. 903 - 913

A Results From Applying Different Levels of Threshold

Classification Output Results

Disease Group	APP	DIV	PPU	NSAP	CHO	INO	PAN	RCO	DYS	Totals
Input Vector										
Sample for APP	284	0	9	81	0	0	4	2	5	385
Sample for DIV	6	23	7	38	1	11	4	0	2	92
Sample for PPU	2	0	42	2	4	1	3	0	2	56
Sample for NSAP	124	24	12	1,422	18	62	14	46	42	1,764
Sample for CHO	6	1	9	39	150	14	22	3	64	308
Sample for INO	14	2	11	36	4	125	18	6	12	228
Sample for PAN	2	1	10	5	3	2	20	3	26	72
Sample for RCO	10	7	1	61	8	4	5	191	0	287
Sample for DYS	6	5	5	104	14	13	28	5	333	513
Totals	454	63	106	1,788	203	232	118	256	486	3,705

Table A-1. The $c^2_{.9}$ Test Data Set Results 69.9%.

Classification Output Results

Disease Group	APP	DIV	PPU	NSAP	CHO	INO	PAN	RCO	DYS	Totals
Input Vector										
Sample for APP	284	0	9	81	0	0	4	2	5	385
Sample for DIV	6	22	7	39	1	11	4	0	2	92
Sample for PPU	2	0	42	2	4	1	3	0	2	56
Sample for NSAP	124	22	12	1,422	19	63	13	46	43	1,764
Sample for CHO	6	1	9	39	149	14	25	3	62	308
Sample for INO	14	3	11	36	4	124	18	6	12	228
Sample for PAN	2	1	12	5	3	2	18	3	26	72
Sample for RCO	10	7	1	58	8	4	6	193	0	287
Sample for DYS	6	6	6	104	14	13	30	5	329	513
Totals	454	62	109	1,786	202	232	121	258	481	3,705

Table A-2. The $c^2_{.95}$ Test Data Set Results 69.7%.

Classification Output Results

Disease Group	APP	DIV	PPU	NSAP	CHO	INO	PAN	RCO	DYS	Totals
Input Vector										
Sample for APP	2,854	0	9	80	0	0	4	2	5	385
Sample for DIV	6	25	7	37	1	10	4	0	2	92
Sample for PPU	2	0	42	2	4	1	3	0	2	56
Sample for NSAP	120	21	13	1,426	19	61	13	49	42	1,764
Sample for CHO	6	1	9	42	148	14	24	3	61	308
Sample for INO	14	2	11	36	4	125	18	6	12	228
Sample for PAN	2	1	11	7	3	2	18	3	26	72
Sample for RCO	11	8	1	54	8	4	5	196	0	287
Sample for DYS	6	6	6	105	14	11	29	5	331	513
Totals	452	64	109	1,789	201	228	118	263	481	3,705

Table A-3. The $c^2_{.975}$ Test Data Set Results 70.1%.

Classification Output Results

Disease Group	APP	DIV	PPU	NSAP	CHO	INO	PAN	RCO	DYS	Totals
Input Vector										
Sample for APP	285	0	9	80	0	0	4	2	5	385
Sample for DIV	5	25	7	37	1	9	6	0	2	92
Sample for PPU	2	0	42	2	4	1	3	0	2	56
Sample for NSAP	119	21	16	1,428	17	59	14	49	41	1,764
Sample for CHO	6	1	9	42	149	14	23	3	61	308
Sample for INO	14	3	11	37	4	121	21	6	11	228
Sample for PAN	2	1	10	7	3	2	20	2	25	72
Sample for RCO	11	8	1	54	8	4	5	196	0	287
Sample for DYS	6	6	5	105	14	11	31	4	331	513
Totals	450	65	110	1,792	200	221	127	262	478	3,705

Table A-4. The $c^2_{.995}$ Test Data Set Results 70.1%.

Classification Output Results

Disease Group	APP	DIV	PPU	NSAP	CHO	INO	PAN	RCO	DYS	Totals
Input Vector										
Sample for APP	285	0	9	80	0	0	4	2	5	385
Sample for DIV	6	25	7	37	1	10	4	0	2	92
Sample for PPU	2	0	42	2	4	1	3	0	2	56
Sample for NSAP	119	21	13	1,426	19	62	13	49	42	1,764
Sample for CHO	6	1	10	42	147	14	24	3	61	308
Sample for INO	14	2	11	36	4	125	18	6	12	228
Sample for PAN	2	1	11	7	4	2	18	2	25	72
Sample for RCO	11	8	1	54	8	4	5	196	0	287
Sample for DYS	6	6	6	105	14	11	29	5	331	513
Totals	451	64	110	1,789	201	229	118	263	480	3,705

Table A-5. The $c^2_{.99}$ Test Data Set Results 70.04%.

B Diagnostic and Symptom Codes for CADA Database

	Symptom	Value
1	SEX	male(1/0), female(1/1)
2	AGE	0-9(2/0), 10-19(2/1), 20-29(2/2), 30-39(2/3), 40-49(2/4), 50-59(2/5), 60-69(2/6), 70 +(2/7)
3	Pain-site Onset	right upper quadrant(3/0), left upper quadrant(3/1), right lower quadrant(3/2),left lower quadrant(3/3), upper half(3/4), lower half(3/5), right half(3/6), left half(3/7), central(3/8), general(3/9), right loin(3/10), left loin(3/11), epigastric(3/12)
4	Pain-site Present	right upper quadrant(4/0), left upper quadrant(4/1), right lower quadrant(4/2),left lower quadrant(4/3), upper half(4/4), lower half(4/5), right half(4/6), left half(4/7), central(4/8), general(4/9), right loin(4/10), left loin(4/11), epigastric(4/12)
5	Aggravating Factors	movement(5/0), coughing(5/1), inspiration(5/2), food(5/3), other(5/4), nil(5/5)
6	Relieving Factors	lying still(6/0), vomiting(6/1), antacids(6/2), milk/food(6/3), other(6/4), nil(6/5)
7	Progress of Pain	getting better(7/0), no change(7/1), getting worse(7/2)
8	Duration of Pain	under 12 hours(8/0), 12-24 hours(8/1), 24-48 hours(8/2), over 48 hours(8/3)
9	Type of Pain	steady(9/0), intermittent(9/1), colicky(9/2), sharp(9/3)
10	Severity of Pain	moderate(10/0), severe(10/1)
11	Nausea	nausea present(11/0), no nausea(11/1)
12	Vomiting	present(12/0), no vomiting(12/1)
13	Anorexia	present(13/0), normal appetite(13/1)
14	Indigestion	history of dyspepsia(14/0), no history of dyspepsia(14/1)
15	Jaundice	history of jaundice(15/0), no history of jaundice(15/1)
16	Bowel habit	no change(16/0), constipated(16/1), diarrhoea(16/2), blood(16/3), mucus(16/4)
17	Micturition	normal(17/0), frequent(17/1), dysuria(17/2), haematuria(17/3), dark urine(17/4)
18	Previous Pain	similar pain before(18/0), no similar pain before(18/1)
19	Previous surgery	yes(19/0), none(19/1)
20	Drugs	being taken(20/1), not being taken(20/1)
21	Mood	normal(21/0), distressed(21/1), anxious(21/2)
22	Colour	normal(22/0), pale(22/1), flushed(22/2), jaundiced(22/3), cyanosed(22/4)
23	Abdominal Movement	normal(23/0), poor/nil(23/1), visible peristalsis(23/2)
24	Abdominal scar	present(24/0), absent(24/1)
25	Abdominal Distension	present(25/0), absent(25/1)
26	Site of Tenderness	right upper quadrant(26/0), left upper quadrant(26/1), right lower quadrant(26/2),left lower quadrant(26/3), upper half(26/4), lower half(26/5), right half(26/6), left half(26/7), central(26/8), general(26/9), right loin(26/10), left loin(26/11), epigastric(26/12), none(26/13)
27	Rebound	present(27/0), absent(27/1)
28	Guarding	present(28/0), absent(28/1)
29	Rigidity	present(29/0), absent(29/1)
30	Abdominal Masses	present(30/0), absent(30/1)
31	Murphy's test	positive(31/0), negative(31/1)
32	Bowel sounds	normal(32/0), decreased/absent(32/1), increased(32/2)
33	Rectal Examination	tender left side(33/0), tender right side(33/1), generally tender(33/2), mass felt(33/3), normal(33/4)

		Diagnostic Groups
	Disease	**Value**
34	APP	Appendicitis
35	DIV	Diverticulitis
36	PPU	Perforated Peptic Ulcer
37	NSAP	Non Specific Abdominal Pain
38	CHO	Cholecystititis
39	INO	Intestinal Obstruction
40	PAN	Pancreatitis
41	RCO	Renal Colic
42	DYS	Dyspepsia