Chapter 7
DIAGNOSIS OF AUTISTIC DISORDER

7.1 The Clinical Decision Making Problem: Diagnosis of Autistic Disorder

Autistic Disorder is a severe developmental disorder, which first becomes evident in early childhood, has lifelong detrimental effects upon an individual's functioning and seems to be closely related, at least phenomenologically, to a number of other identified disorders: Aspergers Disorder; Rhett's Disorder; Childhood Disintegrative Disorder and Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS). Collectively these disorders are referred to as the Autism Spectrum Disorders.

Accurate diagnosis of Autistic Disorder and other Autism Spectrum Disorders, whether for clinical or research purposes can be challenging and it is generally practiced only by clinical professionals who have specialised expertise and experience.

Autistic Disorder and other Autism Spectrum Disorders are rare. Estimates of the prevalence rates for Autistic Disorder and other Autism Spectrum Disorders, in the general population, have ranged from about 5 per 10,000 and 15 per 10,000 respectively [Gray & Tonge, 2001] to 16.8 per 10,000 and 45 per 10,000 respectively [Chakrabarti & Fombonne, 2001].

Autistic Disorder is associated with Intellectual Disability. The rate of intellectual disability in Autistic Disorder is 75-80% [Gray & Tonge, 2001], which itself has a population prevalence rate of 2-3% [American Psychiatric Association, 1994]. From a clinical point of view it is important that children who have an Autistic Disorder are identified from amongst children who have developmental problems (including other Autism Spectrum Disorders) and/or intellectual disability. If identified early, by diagnosis, a child with Autistic Disorder can benefit by receiving autism-specific early intervention. Another important benefit of diagnosis is that families can receive autism specific information and advice, which helps them to assist their child [Gray & Tonge 2001].

The DSM-IV [American Psychiatric Association, 1994] and ICD 10 are the current internationally accepted definitions of the specific behavioural and diagnostic characteristics of Autistic Disorder. The application of these criteria by an experienced clinician in the context of a comprehensive clinical assessment represents the "gold standard" for a diagnosis of Autistic Disorder.

Unfortunately the clinical expertise required to make such diagnoses is not always available and under-diagnosis is a recognised problem [Chakrabarti & Fombonne, 2001, Gray & Tonge 2001]. Diagnostic experts are scarce and when they are available they are relatively expensive. This has led to the investigation of other screening diagnostic techniques based on the use of structured interviews, checklists, or other structured information gathering methodologies. These techniques are used to derive a quantitative

score(s) that can be used as an index of autism. With an appropriate cutoff value the score can be used to provide a criteria for a likely diagnosis of autism. Examples of such instruments are the Autism Behaviour Checklist (ABC), [Krug, Alrick and Almond, 1980], the Autism Descriptors Checklist (ADC), [Friedman, Wolf, Cohen and Fisch, 1985], the Childhood Autism Rating Scale (CARS), [Schopler, Reicher, DeViellis and Daly, 1980] and the Autism Diagnostic Interview-Revised (ADI-R), [Lord, Rutter and Le Couter, 1994]. The obvious advantages of some of these instruments are that they can be completed by those with considerably less expertise (but greater availability), than an expert clinician experienced in the diagnosis of autism. As well if a checklist has been developed according to psychometric principles (Ley 1972), it has known, and usually adequate, reliability and validity.

If structured techniques could be used to make diagnostic decisions that strongly agreed with those of the gold standard (DSM-IV - ICD-10 criteria), then such techniques would be useful tools for clinicians and researchers. Experienced clinicians could use these tools as an independent diagnostic confirmation. Less experienced (with autism) clinicians could use them as a basis for screening cases to be referred on to an appropriate specialist service for diagnosis. Researchers could use such instruments in larger scale population studies to objectively classify subjects without resorting to the use of the expensive “gold standard” clinical diagnosis.

The applicability of neural networks to the diagnosis of Autistic Disorder has been investigated by Cohen et al. [1993] and Cicchetti et al. [1995]. Cohen et al. used data

gathered from a structured parental interview, the Autism Behaviour Interview [Cohen et al 1993] as the basis for classification. They found that a neural network could more accurately classify cases as autistic or non-autistic than the more traditional statistical approach of linear discriminant analysis, using the same input data. The neural network was able to classify correctly 92% of 138 cases (of whom 50% were autistic and 50% were matched controls) when corrected for generalisation error by a "leave 5 out" cross-validation procedure. By comparison the linear discriminant analysis assessed under these same conditions achieved an 82% rate of correct classifications.

Cicchetti et al. [1995] used a sample of 976 cases (454 autistic and 523 non-autistic), which they divided evenly into a training dataset and a cross-validation test dataset. The predictor variables were fifteen ICD-10 diagnostic criteria scored by experienced clinicians and the criterion diagnosis was an "overall clinician diagnosis" (p.28), made by an experienced clinician. They compared three neural networks (two, eight & nine hidden units) with Linear Discriminant Function Analysis (LDFA), Quadratic Discriminant Function Analysis (QDFA) and Logistic Regression (LR) for making the criterion diagnosis of autism. They found that, across several measures of classification accuracy, the statistical techniques (LDFA, QDFA and LR) had either greater cross-validation accuracy and/or less shrinkage than the neural networks. Thus they conclude that the statistical techniques are better.

The studies reported in this chapter have one major important difference to the previous studies, by Cohen et al. [1993] and Cichetti et al. [1995]. In the present studies we use a

parent/carer completed checklist, the Developmental Behaviour Checklist (DBC), [Einfeld & Tonge, 1991,1994,1995], for data gathering in place of the clinician completed structured interviews used by Cohen et al. [1993] and Cicchetti et al [1995].

This has two important implications. First it excludes any possibility of clinician bias affecting the recording of responses. This is important because the "gold standard" for diagnosis, used by Cohen et al. [1993] and Cicchetti et al. [1995] was a clinician made diagnosis, whilst classification in both studies was based on data from structured interviews administered by clinicians. If the same clinician both makes the diagnosis and administers the structured interview, regardless of the order, then there is potential for some form of cross-contamination bias. If, as in the present study, a parent or carer scores the symptom information used for classification and a clinician blind to the parent/carer responses on the symptom checklist makes the criterion diagnosis, then there is no avenue for the predictors and the criterion to become cross contaminated.

A second advantage of the use of a parent/carer-completed checklist is that it is inherently more economical than a clinician-administered structured interview. It requires far less time involvement by a clinician and no clinical expertise to complete. If such a checklist can be shown to achieve a better (or even equivalent) accuracy of diagnosis of autism, than a structured clinical interview, then it is clearly more economical. It may be of value in screening populations or in providing an independent second opinion to clinicians making a clinical diagnosis.

7.2 Study: Comparison of LD and MLP as Classifiers for the Diagnosis of Autistic Disorder using parent/carer responses to the Developmental Behaviour Checklist (DBC)

The objective of this study is to compare linear classification (LD) with non-linear classification (MLPs) for the diagnosis of Autistic Disorder, with the basis for classification being parental responses to the Developmental Behaviour Checklist (DBC).

This will test the hypothesis that the Bayesian decision boundary between cases with Autistic Disorder and Controls, in a decision space defined by the DBC item responses is a non-linear decision boundary.

Specifically this study compares three diagnostic classifiers: a Logistic Discriminant (LD); and two Multi-Layer Perceptron (MLP) Neural Networks with 2 and 3 hidden units. If the Bayesian decision boundary is linear then LD will be the best classifier. If on the other hand the Bayesian decision boundary is non-linear then it is expected that one of the MLP classifiers will be the best classifier and it will classify better than the linear classifier (LD).

Inputs to the diagnostic classifiers consist of a set of parent/carer-completed ratings of behaviour and 3 demographic variables (Age, Sex and IQ range). The criterion diagnosis, which defines the two classes (Autistic Disorder and Control), is a clinician made diagnosis of Autistic Disorder, using DSM-IV criteria.

Subjects and Criterion Diagnosis of Autistic Disorder

Data was obtained from Six hundred and thirty eight (638) children and adolescents, 319 of whom met DSM-IV criteria for Autistic Disorder [American Psychiatric Association, 1994]. The diagnosis of Autistic Disorder was made in the context of five autism assessment services in Victoria and New South Wales, Australia, including the Monash Autism Clinic and four clinics associated with it. These clinics undertake the initial diagnosis of autism for most children, if not all children, with autism in the geographic catchments of the clinics. Diagnosis was undertaken by experienced clinicians working in a multidisciplinary team using a structured combination of observation of the child, information in clinical files, preschool and school reports, direct observation and interview with parents. The other 319 consisted of persons attending general community services for the intellectually disabled and who were assessed as not having Autistic Disorder. All the subjects in the study were diagnosed by either of two experienced senior clinicians (Professor Bruce Tonge, a Child Psychiatrist and Avril Brereton, a Clinical Psychologist). Cohen's kappa was calculated for 52 cases independently diagnosed by both clinicians and was found to be 0.96 indicating a very high level of agreement [Tonge et al 1999, Brereton et al 2002]. The two groups were matched on age (within 2 years), sex and IQ range (4 levels: severe, moderate, mild, borderline/normal) using DSM-IV Mental Retardation range criteria [American Psychiatric Association, 1994].

		Age				IQ Group			
N	M:F Ratio	Mean Age	3 to 6 yrs	7-12 yrs	13 – 19 yrs	Severe	Moderate	Mild	Borderline /Normal
638	502:136	8.9	201	342	95	78	262	188	110

Table 7.1 Sex and Age and IQ group characteristics of the Autistic Disorder Diagnosis Study sample.

Instruments

The Developmental Behaviour Checklist (DBC) [Einfeld & Tonge, 1991,1994, 1995] is a reliable and valid, 96-item checklist designed to assess a broad range of behavioural and emotional disturbances in children and adolescents with mental retardation. Psychometric properties of the DBC are presented in Einfeld & Tonge [1994,1995] and summarised in Appendix A. Clinicians assessing the DSM-IV criteria for Autistic Disorder [American Psychiatric Association, 1994] were blind to parental responses on the Developmental Behaviour Checklist at the time of making their diagnosis.

Input Variable Selection

Each of the 96 items of the DBC were individually examined, using t-tests, to identify those items on which the scores of the Autistic Disorder subjects were significantly different (at a level of $\alpha = .01$), from those of Non-Autistic Disorder controls. Using this procedure 40 of the 96 DBC items were selected and along with age sex and IQ range, making a total set of 43 input variables, used for classification. The input variable set is presented in Table 7.2 below.

Appears depressed, downcast or unhappy
Avoids eye contact. Won't look you straight in the eye
Aloof, in his/her own world
Arranges objects or routine in a strict order
Covers ears or is distressed when hears particular sounds
Doesn't show affection
Doesn't respond to others feelings e.g. shows no response
if a family member is crying
Easily Led by others
Eats non-food items e.g. dirt, grass, soap
Excessively distressed if separated from familiar person
Fears particular things or situations e.g. the dark or insects
Facial Twitches or grimaces
Flicks, taps, twirls objects repeatedly
Fussy eater or has food fads
Gets obsessed with an idea or activity
Has temper tantrums, e.g. stamps feet, slams doors
Hums, whines, grunts, squeals or makes other non-speech noises
Laughs or giggles for no obvious reason
Likes to hold or play with an unusual object, e.g. string, twigs;
overly fascinated with something e.g. water
Mood changes rapidly for no apparent reason
Prefers to do things on his/her own. Tends to be a loner
Preoccupied with only one or two particular interests
Repeated movements of hands, body, head or face e.g. handflapping or rocking
Resists being cuddled, touched or held
Repeats back what others say like an echo
Repeats the same word or phrase over and over
Smells, tastes, or licks objects
Sleeps too little. Disrupted sleep
Stares at lights or spinning objects
Speaks in whispers, high pitched voice, or other unusual tone or rhythm
Switches lights on and off, pours water over and over; or similar repetitive activity
Says he/she can do things that he/she is not capable of
Stands too close to others
Tense, anxious, worried
Unrealistically happy or elated
Unusual Body movements, posture, or way of walking
Upset and distressed over small changes in routine or environment
Wanders aimlessly
Overall, do you feel your child has problems with feelings or behaviour,
in addition to problems with development
Age (in years)
IQ Range (1 = borderline, 2=mild, 3=moderate, 4=severe/profound)
Sex (1=male, 2 = female)

Table 7.2 Input variable set: The first 40 rows are the 40 DBC items, to which a parent responds 0,1 or 2 (0 = not true, 1 = somewhat or sometimes true, 2 = very true or often true) in describing the child's behaviour "now or within the past six months", The last three rows are the three demographic variables Age, IQ Range and Sex.

The 40 DBC items, selected on statistical criteria, also have good face validity as discriminating variables for Autistic Disorder, as they describe many of the behaviours associated with Autism.

Classifiers

1. Logistic Discriminant

An MLP with no hidden units is effectively a Logistic Discriminant (LD). That is the MLP, with no hidden units, approximates the Bayesian classification decision boundary using a linear model. The parameters of that model can be estimated using any appropriate optimisation algorithm. In our case we used the QuickProp algorithm. This is functionally equivalent to carrying out a Logistic Regression (see Chapter 2). The LD was trained using the following parameters: early stopping with a 25% holdout; Cross-Entropy Error Function; Logistic Activation Functions; QuickProp Optimisation: Z score standardisation of all input variables. Bootstrapping was used to obtain a bootstrap corrected A_Z (Area under the ROC Curve) for each MLP network. One hundred (100) bootstraps were used.

2. MLP Neural Network

We trained an MLP type Neural Network, with 3 hidden units. MLPs with a greater number of hidden units where not considered, on the basis that their subject to parameter ratio was far below our minimum model restriction criterion value of 5 (see Chapter 4). The MLP was trained using the following parameters: early stopping with a 25% holdout; Cross-Entropy Error Function; Logistic Activation Functions; QuickProp Optimisation: Z

score standardisation of all input variables. Bootstrapping was used to obtain a bootstrap corrected A_Z for each MLP network. One hundred (100) bootstraps were used.

MLP Model selection

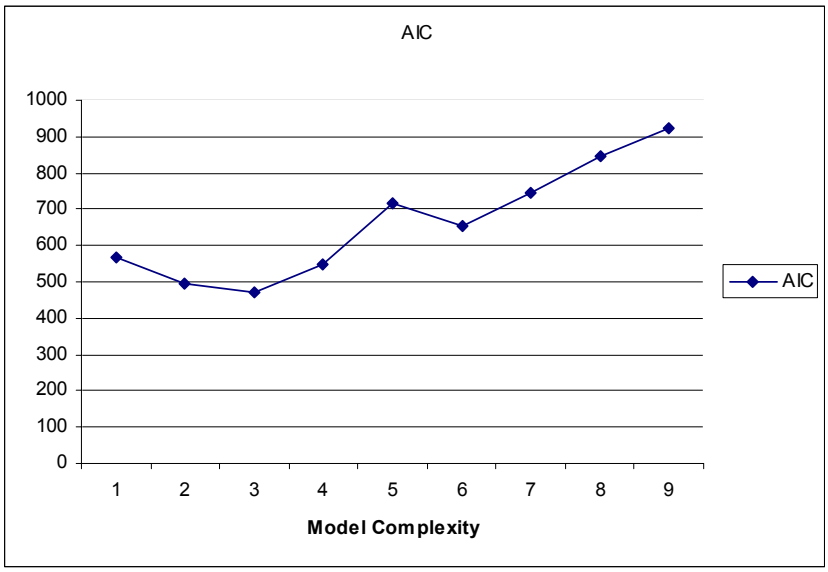

Figure 7.1 AIC values for the LD model (1) and eight MLP models of increasing complexity with 2 to 9 hidden units.

Model	1 (LD)	2	3	4	5	6	7	8	9
Average Cross Entropy	0.38	0.25	0.16	0.15	0.21	0.09	0.09	0.1	0.09
Number of parameters	44	91	136	181	226	271	316	361	406
AIC	**565**	**496**	**473**	**550**	**716**	**655**	**745**	**848**	**925**

Table 7.3. Average Cross Entropy, number of parameters and AIC values for the LD model (1) and eight MLP models of increasing complexity with 2 to 9 hidden units.

From Figure 7.1 and Table 7.3 above the MLP 3 model is selected from amongst the eight MLP models are the comparator MLP model.

Results

	Train Set		Bootstrap		Shrinkage
Model	A_Z	**Error %**	A_Z	**Std Dev**	A_Z
0 LD	.911	15.4	.878	.010	.033
3 MLP	.979	5.0	.931	.011	.048

Table 7.4 Training Dataset and Bootstrap measures of classification accuracy for Logistic Discriminant and MLP with 3 hidden units.

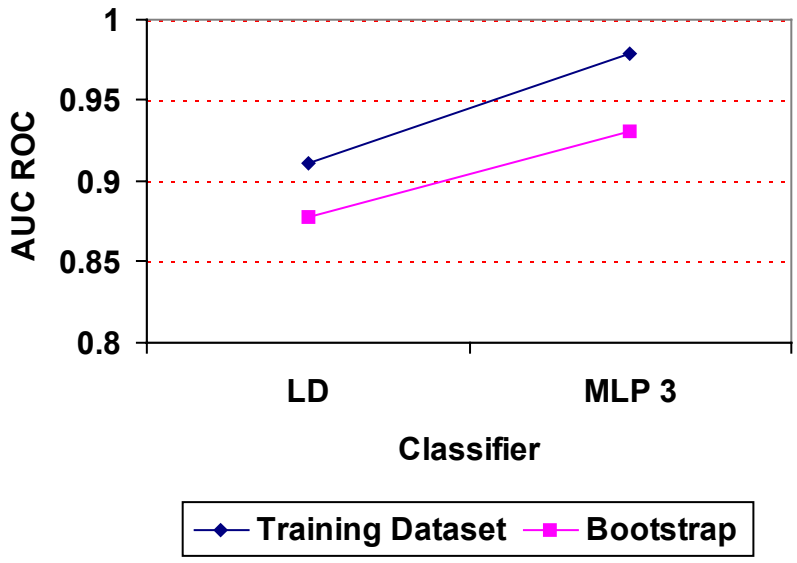

Figure 7.2 Training Dataset and Bootstrap Corrected A_Z (Areas Under the ROC Curve) for LD and MLP with 3 hidden units. Data points are from Table 7.3.

Shrinkage

Comparison of the results in Table 7.4 and Figure 7.2, indicate that both classifiers produced overly optimistic Training Dataset A_Z estimates. In all cases the Bootstrap cross-validation A_Z estimates give lower values

The fact that there is a gap of .03 to .05 A_Z units indicates that the development of classifiers using even larger training dataset sample sizes may produce classifiers with even better accuracy and generalisation.

Classifier Comparison

The best classifier (using bootstrap A_Z as the criterion) was an MLP with 3 hidden units. Using Hanley & MacNeil's [1983] significance test, the difference between the Bootstrap A_Z for the Logistic Discriminant and the MLP Neural Network with 3 hidden units (.878 Vs .931) was significant ($z = 4.3$, $p = .000$, $r_{pos} = .52$, $r_{neg} = .70$).

The training dataset ROC plots (Figure 7.3) demonstrate that the MLP 3 classifies the Training dataset better than the Logistic Discriminant.

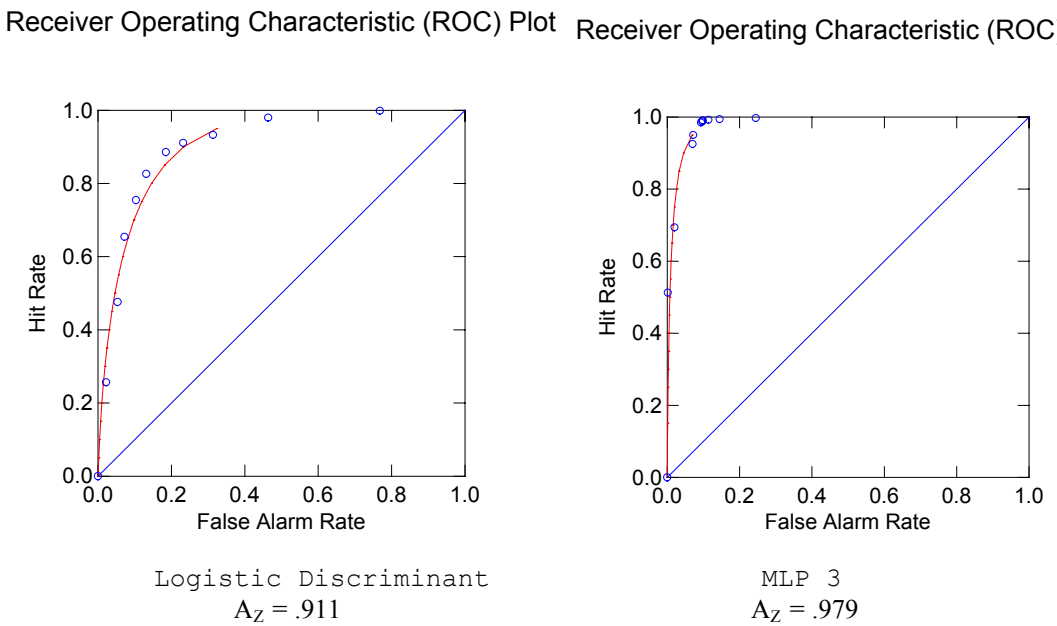

Figure 7.3 Training dataset ROC Curve plots for Logistic Discriminant (LD) and the MLP with 3 Hidden Units.

7.3 Study: Independent Validation of the Neural Network Classifier for Diagnosis of Autistic Disorder

The dataset (N = 638) used to develop an MLP classifier for the Diagnosis of Autistic Disorder in the previous study (Study 7.2) was obtained through the Monash Autism Clinic and collected from clinics in Melbourne, rural Victoria and southern rural NSW. Furthermore only two clinicians (albeit very experienced clinicians) made the gold standard diagnoses of Autistic Disorder or non-Autism for all cases in this dataset.

The follow on study described in this section is based on a new and totally separate data collection carried out in three separate Developmental Assessment Clinics in Sydney NSW. The purposes of doing this were:

- To provide an independent validation test dataset.
- To extend the generalisability of the gold standard by including a new set of clinicians, who are geographically remote from the gold standard clinicians used in study 1.
- To collect a more detailed set of information on diagnoses by asking clinicians to record which DSM-IV criteria for Autistic Disorder each subject met, so that the performance of the Neural Network can be examined in relation to a subject's location on the Autism Spectrum. That is to investigate how the neural network will classify cases, which have some autistic symptomatology, but are classified as

having a Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS), rather than an Autistic Disorder.

The main hypothesis investigated in this study is that the previously developed MLP model for the diagnosis of Autistic Disorder will show good classification accuracy when applied to a totally independent Test Set

Subjects and criterion diagnoses of Autistic Disorder

Three diagnosis and assessment clinics in Sydney, Australia (Grosvenor Diagnosis & Assessment Clinic, Tumbatin Diagnosis & Assessment Clinic and Kogarah Diagnosis & Assessment Clinic) were recruited to obtain an independent validation dataset. Experienced Developmental Paediatricians and Clinical Psychologists staff these clinics. Their primary function is the assessment of children and adolescents with developmental problems. The standard of diagnosis of Autistic Disorder is of the same level as that obtained at the Monash Autism Clinic where the dataset for Study 7.2. Across the three clinics, eight clinicians (4 Developmental Paediatricians and 4 Clinical Psychologists) made DSM-IV diagnoses and completed a DSM-IV Autistic Disorder symptom checklists (See Appendix 1) on each subject.

From the three clinics, a sample of 100 was obtained over a two-year period (1999 - 2000). Characteristics of the sample are presented below:

Source	N	M:F Ratio	Age		
			3 to 6 yrs	7-12 yrs	13 – 19 yrs
Grosvenor	64	55:9	45	11	8
Tumbatin	19	13:6	10	9	-
Kogarah	17	14:3	10	5	2
Total	100	82:18	65	25	10

Table 7.5 Sex and Age characteristics of the Sydney Independent Test Set by clinic source.

Instruments

Parents were asked to complete the DBC checklist as part of general developmental assessment of their child. Clinicians were asked to complete a DSM-IV symptom checklist (see appendix 1) as part of their assessment of the child and prior to viewing any of the responses of the parents to the DBC.

Procedures

The 100 cases in the Sydney Validation dataset were classified by the MLP 3 classifier, which was developed in the previous study (Study 7.3). This assigns a Neural Network Autistic Disorder Probability Score (NNADPS) to each case, which is the probability that the case has a diagnosis of Autistic Disorder. For the purposes of classification, cases with a NNADPS greater than 0.5 are given a neural network diagnosis of Autistic Disorder and

those with a NNADPS of 0 to 0.5 are given a classified as not having a diagnosis of Autistic Disorder.

Results

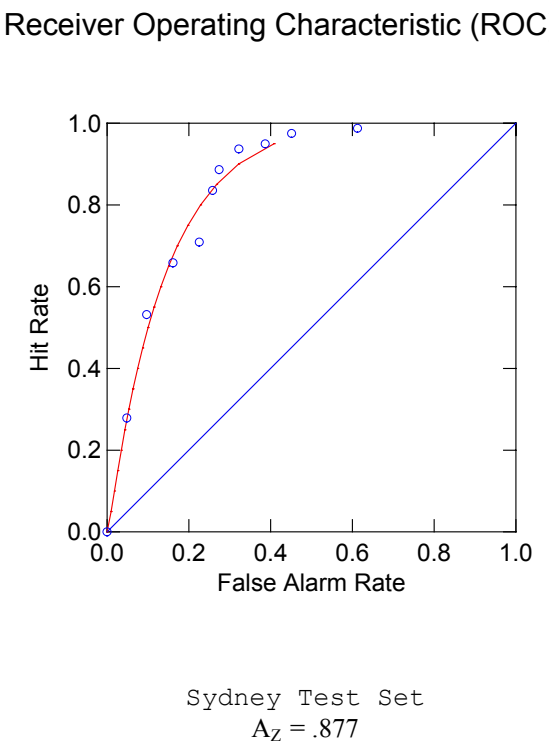

Figure 7.4 ROC Plot for the Sydney Test Set classified by the MLP 3 developed in Study 7.3.

The ROC plot in the above figure has an Area under the curve of .877. This indicates the MLP 3 classifier is a "Good" classifier of this dataset's previously unseen cases, collected at similar but geographically remote clinics, which are inturn independent of the Monash Clinic from which the Training dataset was collected.

Agreement between Neural Network diagnoses and clinician's "Gold Standard" DSM-IV are presented in the Table below.

	Neural Network Diagnosis		
Gold Standard Diagnosis	**Autistic Disorder**	**Non Autistic Disorder**	
DSM-IV Autistic Dis.	35	3	38
DSM-IV Non Autistic	17	45	62
	52	48	100
Overall Accuracy	80%	(95% CI: 71-87)	
Sensitivity	92%	(95% CI: 79-98)	
Specificity	73%	(95% CI: 60-83)	
A_Z	.877	(95% CI: .80-.95)	

Table 7.6 Classification Accuracy of the Neural Network Diagnosis of Autistic Disorder on the Sydney Independent Validation Test Set.

The accuracy rates across the three clinics were all the same, 80% (Grosvenor 50/64, Kogarah 14/17 and Tumbatin 16/19), and therefore all equal to the overall accuracy rate for the Sydney Test Set as a whole. This suggests that the same Gold Standard diagnoses were being applied at all three clinics. It also suggests that the Sydney Gold Standard is in agreement with the Melbourne Gold Standard.

7.4 Neural Network Autistic Disorder Probability Score Verses Test Set Probability

According to Neural Network Theory (see Chapter 2), the output of our MLP neural network, trained with a cross-entropy error function, can be directly interpreted as the probability that an individual will have a Gold Standard Diagnosis of Autistic Disorder. Thus, if an individual receives a Neural Network Autistic Disorder Probability Score (NNADPS) of 0.3, it can be interpreted as 30% of individuals with the same profile of input scores (on the 43 input variables) will be given a DSM-IV diagnosis of Autistic Disorder, and 70% will not. It follows from this that the empirical distribution of diagnoses in the Test Set will approximate the conditional distributions predicted by multiplying the middle value of the NNADPS (i.e. the probability of a diagnosis) at each interval by the number of cases with a NNADPS value in that interval.

Table 7.7 below tests this hypothesis for the Sydney Test Set: that the empirical distribution of Autistic Disorder cases at different levels of Neural Network Autism Probability Score (NNADPS) approximates the distribution predicted by multiplying the middle value of the NNADPS at each interval by the number of cases with a NNADPS value in that interval.

NNADPS	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
N	29	8	4	5	2	8	3	8	10	23
Expected	2	1	1	2	1	4	2	6	9	22
Actual	0	1	0	2	0	6	1	3	6	19

$©^2 = 8.41$, not significant at 0.05

Table 7.7 Actual and expected (from Neural Network Autistic Disorder Probability Score) distributions of cases diagnosed as Autistic Disorder. The $©^2$ for the differences between the distributions of 8.41 did not exceed the critical value of 16.93 (df=9, ⟨= 0.05)

The empirical distribution of Autistic Disorder cases at each level of NNADPS and the expected distribution predicted by NNADPS are not significantly different from each other ($©^2 = 8.41$, $p > 0.05$).

7.5 Neural Network Autistic Disorder Probability Score Verses the Number of DSM-IV Symptoms

A clinician making a DSM-IV Diagnosis of Autistic Disorder examines twelve criteria and rates each criterion as being present or absent for that individual. If six or more criteria are present, the deficits of autism were present before the age of 3 years, and the symptoms are not better accounted for by either Rett's Disorder or Childhood Disintegrative Disorder, then a diagnosis of DSM-IV Autistic Disorder is made.

If an individual meets some of the criteria for Autistic Disorder, but not enough to meet full criteria for Autistic Disorder, the clinician can make a diagnosis of Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS), which is said to belong to the Autistic Spectrum of Disorders and is seen as a milder form of Autism. If PDD-NOS were a milder form of autism, then it would be expected that individuals with PDD-NOS would occupy the middle ground between Autism and Non-Autism on measures of Autistic symptomatology. As such it might be expected that PDD-NOS individuals will obtain Neural Network Autism Probability Scores in the middle range. As well, if PDD-NOS is a milder form of Autistic Disorder, It might also be expected that there will be a monotonic increase in the number of DSM-IV symptoms observed in relation to increases values of the Neural Network Autistic Disorder Probability Score.

The cross distribution of number of DSM-IV symptoms by Neural Network Autistic Disorder Probability Score, for the Sydney Test Set, is presented in tables 7.8 and 7.9 below. Subjects were classified into diagnostic groups by the number DSM-IV symptoms

scored by clinicians. A subject with 0, 1 or 2 symptoms was classified as Not Autistic, a subject with 3, 4 or 5 symptoms was classified as PDD-NOS and a subject with 6 or more symptoms was classified as having Autistic Disorder (DSM-IV manual, APA [1994]). All the subjects who had more than 5 symptoms also met the other criteria required for a DSM-IV diagnosis of Autistic Disorder.

NN Prob Score	Not Autistic Spectrum			PDD-NOS			DSM-IV Autistic Disorder			
0.9				1	1	2	16	1		2
0.8				1	2	1	4	1	1	
0.7	2		1	1	1		3			
0.6	1			1			1			
0.5		1		1			4	1	1	
0.4	1				1					
0.3		1			2		2			
0.2			1	2	1					
0.1	4	1		1	1			1		
0.0	15	4	4	5	1					
	0	1	2	3	4	5	6	7	8	9+

Number of DSM-IV Autistic Disorder Symptoms

Table 7.8 Cross-distribution of Neural Network Autistics Disorder Probability Score and DSM-IV symptom score for the Sydney Independent Validation Test Set.

Table 7.8 below collapses Table 7.7 above, by reducing the cross distribution of subjects into 4 NN Output Score ranges by a 3 way diagnostic classification of Non-Autistic, PDD-NOS and Autistic Disorder.

	Gold Standard Diagnosis			
NNADPS	**Non Autistic Spectrum**	**PDD-NOS**	**DSM-IV Autistic Disorder**	**Total**
70 – 100	3	10	28	41
51-69	2	2	7	11
31-50	2	3	2	7
0 – 30	29	11	1	41
Total	36	26	38	

Table 7.9 Cross-distribution of Neural Network Autistic Disorder Probability Score and DSM-IV Diagnosis for the Sydney Independent Validation Test Set.

From Table 7.9 we can see that the distributions for the two diagnostic end groups (Non-Autistic and Autistic Disorder) are both unimodal and appropriately skewed to the low and high ends of the Neural Network Autistic Disorder Probability Score range. However the distribution of Neural Network Autistic Disorder Probability Scores for PDD-NOS group is bimodal, with strong skewing away from the centre of the range. This is not consistent with the hypotheses that PDD-NOS represents a mild for of Autistic Disorder

These distribution patterns suggest that, in terms of the input variables used to discriminate the two groups, the Autistic Disorder and Non-Autism Spectrum group are well characterised. However the PDD-NOS group, rather than being an intermediate form between Autistic Disorder and Non-Autism Spectrum, is heterogeneous group composed of at least two subgroups. One subgroup with similarity to Autistic Disorder which may be a milder form of the Autistic Disorder and another subgroup, which is not similar to Autistic Disorder, despite being rated by clinicians as having significant autistic symptomatology. Contrary to expectation, very few of the cases diagnosed as PDD-NOS occupied the middle ground of NNADPS range. The conceptual map presented in Figure 7.5 below, gives one possible explanation for the pattern of results observed in Tables 7.8 and 7.9.

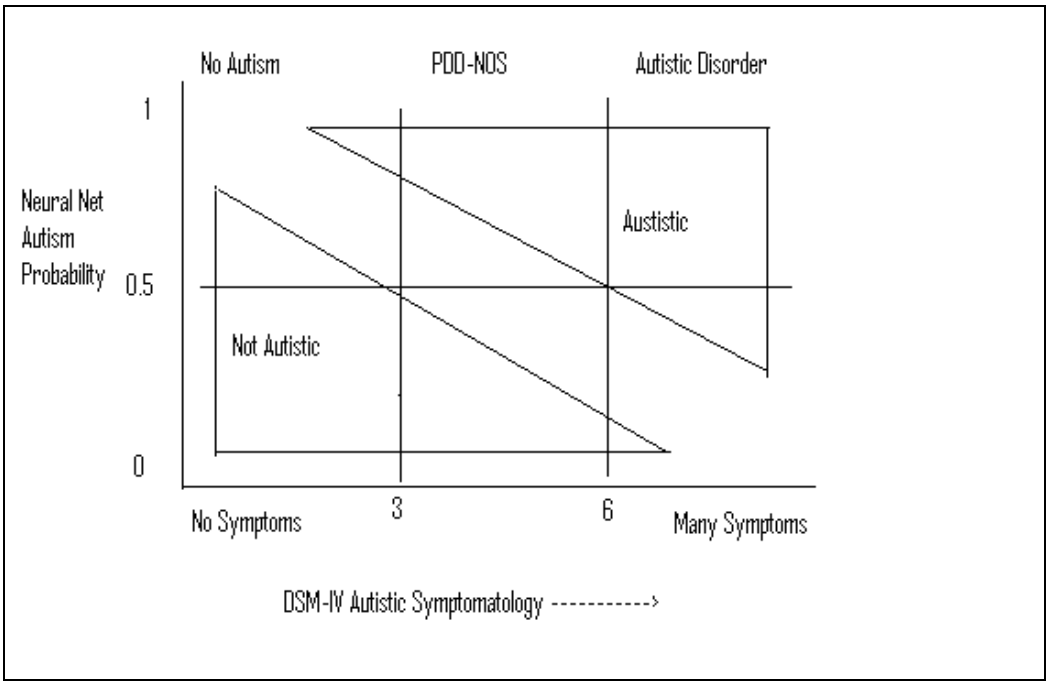

Figure 7.5 Conceptual Map of the distributions of Autistic and Non-Autistic populations in reference to clinician defined DSM-IV Symptomatology and NN Autism Probability derived from parent-reported patterns of behaviour.

7.6 The Stability of Neural Network Diagnosis of Autistic Disorder Over Time

A diagnosis of Autistic Disorder is relatively stable and likely to be of lifelong duration. Thus it is hypothesised that individuals who receive a Neural Network Diagnosis of Autistic Disorder based upon DBC items scores and Age, Sex and IQ range, should again by classified as having a Diagnosis of Autistic Disorder at a much later time, by the Neural Network using a fresh set of input data obtained at that later time.

Subjects

As part of a separate longitudinal epidemiological study, which used the DBC (Tonge and Einfeld, 1998) we had access to DBC data on 40 subjects diagnosed with autism according to DSM-IV criteria. These subjects who were in the age range 4 to 18 years at the commencement of the study (Time 1) had DBC checklists completed by parents or carers at Time 1 and five years later (Time 2).

A diagnosis of Autistic Disorder should be relatively stable over a 5-year period. Thus for individuals diagnosed as having an Autistic Disorder, it could be expected that 5 years later they would still retain the behavioural features of autism and again receive a diagnosis of Autistic Disorder. The availability of data from the aforementioned Longitudinal Study of DBC checklist data at two points in time spread 5 years apart (1991 and 1996) makes it possible to test this hypothesis about the stability of the neural network diagnosis of Autistic Disorder.

Results

	Time 2 (5 years later)		
Time 1	**NN Autistic Disorder**	**NN Non-Autistic**	**Total**
NN Autistic Disorder	36	0	36
NN Non Autistic	3	1	4
Total	39	1	40

Table 7.10 Correspondence between Time 1 and Time 2 (5 years later) neural network diagnoses on the same 40 individuals all diagnosed as DSM-IV Autistic Disorder at Time 1.

Table 7.10. shows that for a group initially clinically diagnosed with Autistic Disorder those diagnosed as having Autistic Disorder by the neural network will again be diagnosed with Autistic Disorder by the neural network using a second DBC checklist completed 5 years later. Of the four individuals initially classified as Non-Autistic Disorder by the neural network, three were later re-classified as having Autistic Disorder 5 years later. Only one individual was classified as Non-Autistic Disorder on both occasions. Thus a neural network diagnosis of autism appears to be highly stable over a 5-year period for individuals who had a clinical diagnosis of Autistic Disorder.

7.7 Conclusions from Studies

The studies in this chapter examined the capacity of MLP type Neural Networks for making a diagnosis of Autistic Disorder

Study 7.2, found that MLPs classified cases, as DSM-IV Autistic Disorder or not, better than did a Logistic Discriminant. This finding supports the hypothesis that the Bayesian decision boundary between DSM-IV Autistic Disorder and Controls is essentially non-linear.

Study 7.3 examined the use of the diagnostic classifier developed in Study 7.2 with a set of 100 new cases obtained from three separate clinics in Sydney, which were geographically remote from the clinics (Victoria and Southern NSW) used to obtain the Training Dataset used in Study 7.2. The level of classification accuracy in the Sydney dataset was similar to that found in the Melbourne Test dataset.

Study 7.4: in the Test Set from Sydney (Study 7.3) the empirical distribution of actual diagnoses of Autistic Disorder by Neural Network Autism Probability Score (NNADPS) was as expected from the distribution of all cases by NNADPS. This validates the accuracy of the NNADPS as a measure of the posterior probability that an individual has a diagnosis of Autism.

In Study 7.5, an examination of the relationship between NNADPS and the number DSM-IV Autistic Disorder symptoms indicated that, as predicted, individuals with zero, one or

two symptoms (no Autistic Spectrum Disorder) generally had a low NNADPS and individuals with 6 or more symptoms generally had a high NNADPS. It was predicted that individuals with three, four or five DSM-IV symptoms (who are diagnosed as Pervasive Developmental Disorder – Not Otherwise Specified [PDD-NOS]), would have a medium NNADPS. However this was not found to be the case. The distribution of NNADPS for PDD-NOS subjects was instead bi-modal, skewing towards both the high and low ends of the NNADPS range. This suggests that the PDD-NOS group is not (contrary to most theories) a mild form of Autistic Disorder, but is more likely to be a heterogeneously composed group, which contains at least two subgroups.

Finally in Study 7.6, it was found that the Neural Network Diagnoses of Autism was very stable over a 5-year period for a group of 40 individuals diagnosed as having an Autistic Disorder.

Collectively these findings indicate that a Neural Network Diagnosis of Autism based upon the parent/carer completed DBC Checklist will be a clinically useful addendum to the toolbox of clinicians involved in making a diagnosis of Autism.

7.8 Comparison to Currently Used Practices for Diagnosis of Autistic Disorder

There are a number of practices, which are currently in wide usage by clinicians for making a Diagnosis of Autistic Disorder.

DSM-IV\ICD-10

DSM-IV and ICD-10 criteria for Autistic Disorder[1] are the "Gold Standard" practice for making a Diagnosis of Autistic Disorder. The criteria are structured into four collections:

1 Qualitative Impairment in Social Interaction (4 criteria)

2 Qualitative Impairment in Communication (4 criteria)

3 Restricted Repetitive & Stereotyped Patterns of Behaviour, Interests and Activities (4 criteria)

4 Delays or Abnormal Functioning, with Onset prior to age 3 years (3 Criteria)

A Diagnosis of Autistic Disorder is made when then number of criteria met out of collections 1,2 and 3 is 6 or more, with at least 2 criteria in collection 1 being met, and in addition there is at least one criteria met from collection 4. A Diagnosis of Autistic Disorder is pre-empted by a Diagnosis of Retts Disorder or a Diagnosis of Childhood Disintegrative Disorder.

[1] DSM-IV and ICD-10 Criteria are functionally identical with respect to making a Diagnosis of Autistic Disorder

Cases where there a number of criteria (more than two), but not sufficient to meet the requirements for a Diagnosis of Autistic Disorder can be given a diagnosis of Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS) under DSM-IV or Atypical Autism under ICD-10. In cases where none of the Communication (collection 2) criteria are met, but some Social Interaction (collection 1) and Cognitive Rigidity (collection 3) criteria are met, and Intellectual Functioning is within the normal range, then a Diagnosis of Aspergers Disorder can be made. Collectively all these disorders are commonly grouped together under the title of Autism Spectrum Disorders. Though, the validity of this "spectrum" concept is under question (Tonge 2002).

In order to evaluate the criteria and thus be able to make a Diagnosis of Autistic Disorder, a clinician needs to be experienced in developmental disorders and needs to collect criteria relevant information via direct observation, parental interview and a developmental history. In many cases a multi-disciplinary team is used to make the diagnosis (Tonge, 2002) and/or information is gathered from a wide set of sources, such as educators in pre-schools and schools, speech pathologists, early childhood health services, general practitioners and other clinicians.

DSM-IV\ICD-10 criteria cannot be validated – because they are the "Gold Standard". But Reliability (agreement) is 65% (other PDD Vs Autism) to 95% (PDD Vs No PDD), Aspergers Vs other PDD is 60%. **For Autistic Disorder Vs non-PDD agreement is 85%.** Therefore, theoretical validity of other methods cannot exceed this level, since they cannot agree better with the "Gold Standard" than the "Gold Standard" does with itself.

Childhood Autism Rating Scale (CARS)

The CARS [Schloper et al 1980] is a clinician rating scale. The clinician rates the child on 17 items. Each item is on a 5-point scale. Each point on each item is anchored with a definition, and 1/2 point scores are possible. The test-retest Pearson correlation is .88. The CARS test manual also reports Pearson correlations of .84 and .80 between CARS score and clinicians' ratings of autism severity. It has an 82% agreement with clinical diagnosis. That is a clinical judgment that a child has Autism rather than a diagnosis made according to some criteria such as DSM-IV.

Gilliam Autism Rating Scale (GARS)

The GARS [Gilliam, 1997] is a parent-completed instrument. It contains 56 items. The GARS test Manual reports 90% agreement between a GARS diagnosis of Autism and Clinical diagnosis made by school district diagnostic personnel, but the diagnostic criteria used is not mentioned.

Autism Diagnostic Observation Schedule (ADOS)

The ADOS [Lord et al 2000] is a structured schedule of observations of the child in response to "presses", contrived situations which act as stimuli for behavioural responses. The clinician delivers the "press" and then rates the child's response. It is not dissimilar to developmental assessment using an instrument such as the Griffiths Mental Development Scales. There are four modules. Each designed for a particular developmental level – a child is assessed using only one module that fits them. It takes about 45 minutes to

administer. Assessors have to be trained and certified. The ADOS has been found to be 95% correct for Autism, 92% correct for non-autism, 33% correct for PDD-NOS (53% were labeled as Autism, 14% as Non-Autism).

Autism Diagnostic Interview – Revised ADI-R

The ADI-R [Lord et al, 1994] is clinician based structured interview. Items are scored on a 0-3 scale. It can take up to 2 hours to administer. It has a diagnostic algorithm based upon DSM-IV to give PDD diagnoses. Based upon DSM-III-R it had 92% accuracy.

Developmental Behaviour Checklist – Autism Screening Algorithm

The DBC-ASA was developed by Brereton et al [2002], using Logistic Regression and Factor analysis to identify 29 DBC items which related to a DSM-IV diagnosis of Autistic Disorder. The DBC-ASA is a unit linear combination of the raw scores. That is the raw scores on the 29 items are added together to get a DBC-ASA score (range 0 – 58). By ROC curve analysis it was determined that the optimal cutoff for a diagnosis of DSM-IV Autistic Disorder was 17. This gave a Sensitivity of 86%, a specificity of 69%, an overall accuracy of 78% and an Area Under the ROC Curve of 0.80.

Neural Network Diagnosis of Autistic Disorder

The Neural Network diagnosis of Autistic Disorder developed earlier in this chapter is also based upon the Developmental Behaviour Checklist (DBC). The training dataset derived agreement between a Neural Network assigned Diagnosis of Autistic Disorder and the Gold Standard was 92%. (Area Under the ROC Curve of 0.98). Test dataset

derived agreement was 80% for the Sydney Test dataset (Area Under the ROC Curve of 0.88).

Autism Diagnostic Instruments - Compared

Method\Practice	Training dataset agreement	Test dataset agreement
DSM-IV	Not Applicable	85% Agreement
CARS	82% Agreement	?
GARS	90% Agreement	?
ADOS	95% Agreement	?
ADI-R	92% Agreement	?
DBC-ASA	78% Agreement	?
DBC-NN	92% Agreement	80 % Agreement

Table 7.11 Training dataset derived and Test dataset derived agreement with Gold Standard (DSM-IV) of commonly used Instruments and Practices used in the Diagnosis of Autistic Disorder and the Neural Network Diagnosis of Autistic Disorder based upon the DBC

From Table 7.11 above it can be seen that the DBC-NN method has training dataset derived agreement that is in the same range as that of any of the other commonly used practices and methods and that its test dataset derived agreement with a DSM-IV/ICD-10 Diagnosis of Autistic Disorder is in the same range as the average agreement of one clinician with another clinician using DSM-IV/ICD-10 criteria.

7.9 Envisaged Clinical Use(s) of the DBC-Neural Network Diagnosis of Autistic Disorder

What role(s) can be envisaged for the DBC-Neural Network Diagnosis of Autistic Disorder (DBC-NN) developed earlier in this chapter?

The DBC-NN is easy to incorporate into existing practices. It makes a demand of 15 to 20 minutes of time upon a parent or carer and about 5 minutes upon a clinician. No explicit training is required, but the clinician needs to be familiar with the DBC and its manual. Data entry of DBC scores and other data onto a computer is straight forward, and calculation of the probability of a Diagnosis of Autistic Disorder is carried out by computer. The clinician would need to know how to interpret the probability.

The DBC-NN unlike most other practices (except the GARS) is totally independent of the clinician. That is no action of the clinician has a causal relationship with the obtained Autistic Disorder probability score. Other practices, in which the clinician participates as data collector, rater or interpreter of criteria, are all open to contamination of the resultant outcome with biases of the clinician. For example, if the clinician has made an implicit clinical judgment that the child has or does not have an Autistic Disorder, before or during the course of making ratings or scoring criteria, then the ratings or scores might be biased to agree with the clinician's judgment.

Using the DBC-NN the clinician, can go ahead and use any of the other practices, arrive at diagnosis and then check the DBC-NN diagnosis as an independent second opinion. In

fact it is recommended, that if a clinician decides to incorporate the DBC-NN into their practices for testing a Diagnosis of Autistic Disorder, they should use their other practices first, in particular DSM-IV\ICD, which is the "Gold Standard" and then use the DBC-NN probability as an independent second opinion, which confirms their diagnosis or causes them to re-examine it.

It is not recommended that the DBC-NN should be used as a primary method of diagnosis. The "Gold Standard" is DSM-IV\ICD-10, therefore it should always be the primary practice for a Diagnosis of Autistic Disorder.

If the DBC-NN is to be used as part of a mix of practices (this is generally a good approach in respect of Diagnosis of Autistic Disorder), and some of the other practices require the clinician to rate or score criteria, then the DBC-NN should be finalized (the final step of obtaining the autism probability) after the other practices have been completed by the clinician. This is recommended to prevent a reverse bias effect where knowledge of the DBC-NN outcome influences the clinician, and biases their ratings or scoring of criteria in such a way as to produce an outcome for these other practices (including DSM-IV/ICD-10), which is different from what it would otherwise have been.

Finally it is not recommended that the DBC-NN be used for screening and further referral. The specificity of the DBC-NN is 73 percent (see table 7.6). Therefore 27 percent of children, without Autistic Disorder "seen" by the DBC-NN will be labeled as having an Autistic Disorder. In screening situations the base rate for the disorder is low (from a few percent to 20 percent) and the majority of the children the DBC-NN "sees" will not have

an Autistic Disorder, only a small minority will. In such a situation the majority of the children "screened" as positive (i.e. labeled by the DBC-NN as having an Autistic Disorder) will not actually have an Autistic Disorder. That is, the Positive Predictive Value of the DBC-NN in a screening situation is very low. See Figure 7.6 below.

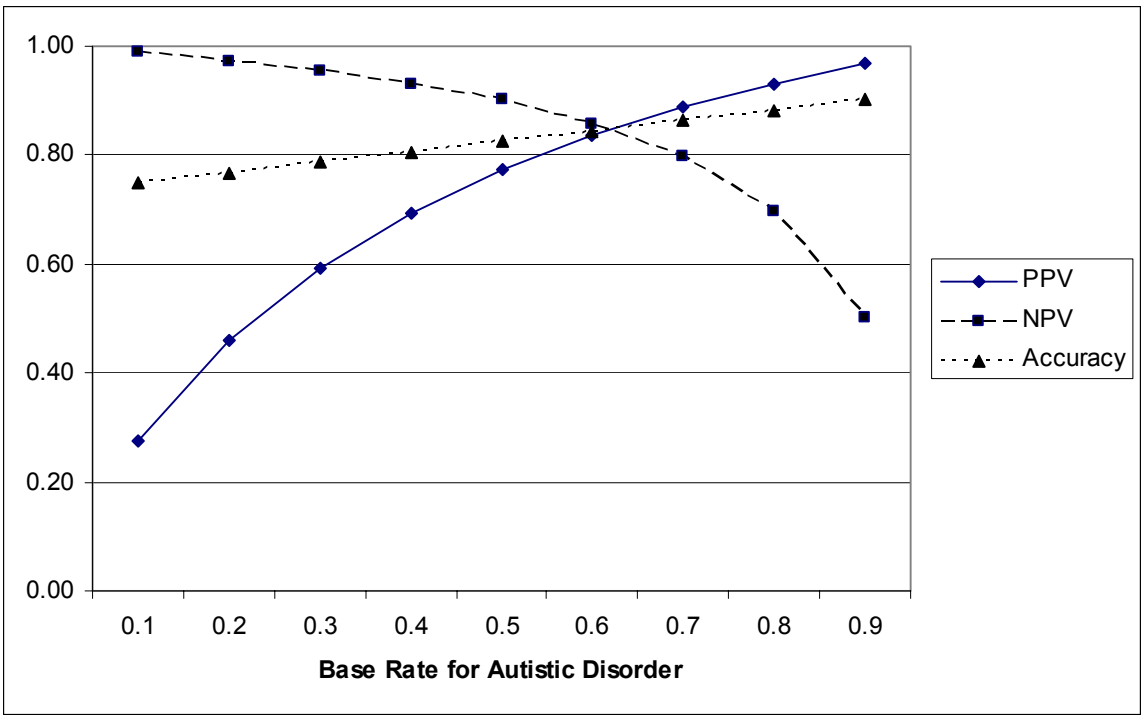

Figure 7.6 The effects of the Base Rate for Autistic Disorder (in the clinical population it is applied to) upon the Positive Predictive Value (PPV), Negative Predictive Value (NPV), and overall Accuracy of the DBC-NN. Calculations are based upon the Sensitivity and Specificity values given in Table 7.4.

In a high base rate situation (70 percent to 90 percent), the converse applies and the Positive Predictive Value of the DBC-NN will be high (see Figure 7.6 above). It is in such a situation that use of the DBC-NN is recommended. The "Second opinion" role suggested earlier is such a situation.