Chapter 6:
PREDICTION OF RESPONSE TO TREATMENT WITH STIMULANT MEDICATION IN CHILDREN WITH ADHD

6.1 The Clinical Decision Making Problem: Prediction of Response to Treatment with Stimulant Medication in Children with Attention Deficit Hyperactivity Disorder (ADHD).

Attention Deficit Hyperactivity Disorder (ADHD) is a common behavioural disorder of childhood and adolescence. ADHD affects 3 to 5% of the school aged population, is about four times more prevalent in males than in females, and it is the most common cause of referral to child and adolescent psychiatric services (APA 1994). It is characterised by (and diagnosed by) abnormal levels of hyperactivity-impulsivity and inattention. A child is diagnosed, according to DSM-IV, as having ADHD if they display six or more primary symptoms in list of 18 symptoms, grouped into Hyperactivity, Impulsivity (9 symptoms) or Inattention (9 symptoms) (APA [1994], American Academy of Child & Adolescent Psychiatry [1997]). The DSM-IV diagnostic criteria for ADHD (APA [1994]), elaborate three basic subtypes: ADHD - Primarily Inattentive (where clinical criteria for Hyperactivity-Impulsivity are not met); ADHD - Primarily Hyperactive-Impulsive (where clinical criteria for Inattention are not met); and the ADHD - Combined Type (where clinical criteria for both Hyperactivity-Impulsivity and for Inattention are met). A fourth category, ADHD - Not Otherwise Specified (NOS) is

sometimes used to describe those with obvious symptoms, but who do not meet criteria for any of the three DSM-IV recognised ADHD subtypes.

ADHD sequelae: disruptiveness, increased demands for supervision by parents, carers and teachers; strained relationships with parents, carers and teachers; poor social functioning with peers; educational under-achievement and poor self esteem [Barkley et al 2002]. Follow-up of children with ADHD into adulthood has indicated that as adults these individuals have higher than expected rates of involvement with criminal justice system, incarceration, drug and alcohol problems, attempted suicide and suicide [Barkley et al 2002]. The primary symptoms (hyperactivity, impulsiveness and inattention) are thought to be associated with secondary immediate and long-term problems outlined above. ADHD in at least some individuals continues into adulthood [Gittleman et al 1985]. Effective treatments of the primary symptoms should result in significant short-term and possible long-term benefits to the affected individual.

The benefits of psycho-stimulant medication for the treatment of ADHD in children are firmly established. In a recent review, Greenhill, Halperin and Aikoff [1999], the pooled results of 5,899 children, participating in 161 randomised controlled trials, found that 65% to 75% of children treated with psycho-stimulant medication showed clinical improvement, whilst the rates for placebo controls ranged only from 4% to 30%. The MTA study [MTA Cooperative Group, 1999], a randomised clinical trial of psycho-stimulant medication, behavioural intervention, and combined treatment found medication was the most effective intervention.

However, there is a public & professional concern that stimulants are being over prescribed to children and adolescents. When she was the US First Lady in 2000, Hillary Clinton publicly expressed her view that stimulant medications were being over-prescribed to American children by US clinicians. In a recently published "Clinical Practice Guideline: Treatment of the School-Aged Child with Attention-Deficit/Hyperactivity Disorder", the American Academy of Paediatrics [2001] highlights that there is a need for quality improvement in both the accuracy of diagnosis and in treatment decision making, by primary care clinicians. In NSW, in 2001, a State Government Senate Committee on Social Issues was set up to examine the use of stimulants with children in NSW. In Australia, stimulants can only be prescribed to children by selected medical specialists, and the prescribing doctor is required to obtain an individual authority for each prescription from the Commonwealth. In New South Wales (and also in Western Australia) there is a 'Stimulants Committee', which reviews psycho-stimulant prescribing practices in that state.

Much of this public and professional concern is based on anecdotes and/or intuitions. Importantly though, there are empirical data which support these concerns. Jensen et al, [1999] found in a four communities epidemiological sample that only 50% of children who had been prescribed stimulants in the preceding 12 months met the full DSM-III-R criteria for ADHD. They also found that whilst 5.1% of children in their community samples met full DSM-III-R criteria for ADHD only 12.5% of these had received treatment with stimulant medication in the preceding 12 months. As Jensen et al [1999]

point out these data raise questions about the possibility of both over-prescription and under-prescription of stimulant medication in children

Over-prescription, the main focus of public concern, is of concern because the inappropriate use of CNS stimulants, for children who are unlikely to receive a clinical benefit, carries with it morbidity in terms of anorexia, headache, dysthymia and precipitation of tics in vulnerable children. Under-prescription, which is a less appreciated problem, is also of concern because children who would obtain a significant benefit are left untreated.

The current trend is for progressively increasing numbers of children to be prescribed psycho-stimulant medication [Greenhill et al, 1999, Zito et al, 2000]. It is not at all clear that these increasing numbers of prescriptions are doing anything to address the problem of under-prescription, and could inadvertently be contributing to increased over-prescription [Jensen, 2000].

An important corollary to the findings of Greenhill et al [1999] is that if about 65% to 75% of children, who have presumably met clinical diagnostic criteria for ADHD, respond, then between 25% and 35% of these children do not show a clinically significant degree of improvement, when treated with a psycho-stimulant medication. It has been postulated that these differences in clinical response to psycho-stimulants are genetically determined [Winsberg & Comings, 1999]. In addition it is an open question, if there also exists, a group of children with some ADHD symptoms who do not meet all the formal

criteria for a diagnosis of ADHD, but who none-the-less clinically benefit from treatment with psycho-stimulant medication.

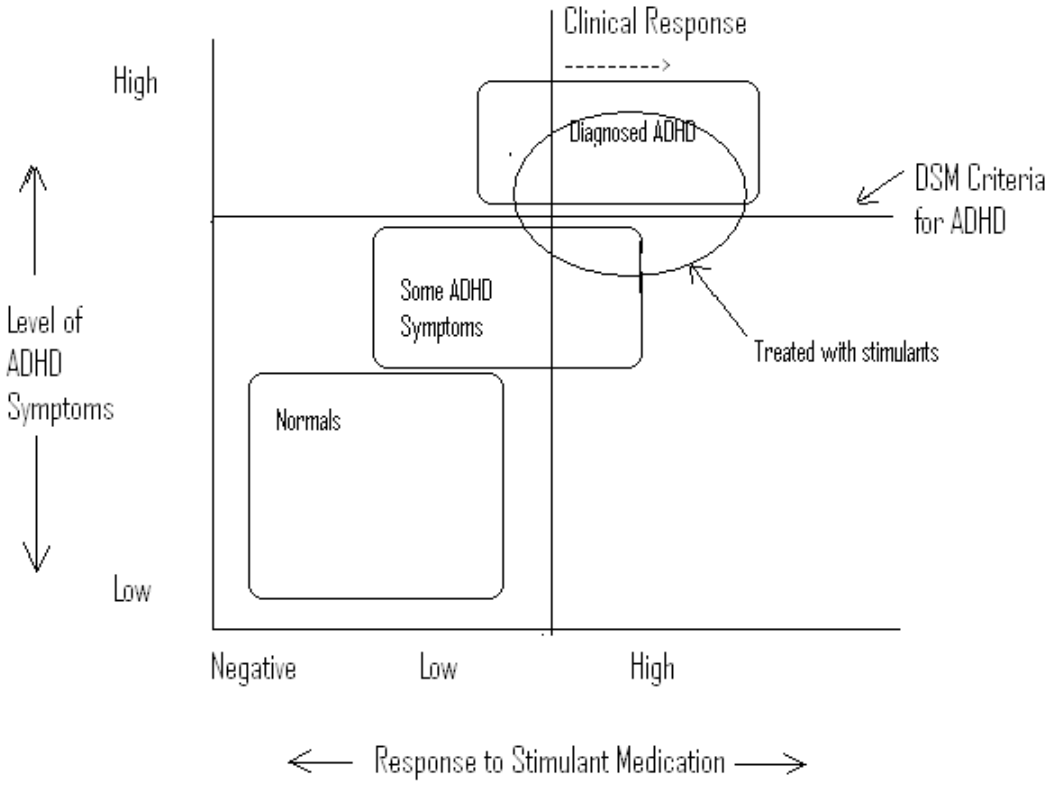

Figure 6.1 Conceptual Map of ADHD symptomatology, potential responsiveness to treatment with Stimulant Medication, and actual treatment provided by clinicians.

In Figure 6.1 the vertical dimension measures the Level of ADHD symptomatology, which ranges from low to high. There are 3 groupings of children: *Normals*, children with none or very few symptoms of ADHD; *Some ADHD Symptoms*, a group of children who

have some symptoms of ADHD but who fall below the criteria required for a DSM-IV diagnosis of ADHD and; *Diagnosed ADHD*, the group of children who have symptoms of ADHD and who also meet all the criteria for a DSM-IV diagnosis of ADHD. The horizontal dimension of the map measures the degree of benefit to children of response to treatment with stimulant medication, which ranges from a negative benefit through a low benefit to a high benefit. At a certain level, the degree of benefit is classed as a clinically beneficial response. What the map indicates is that a clinically beneficial response occurs for a significant majority (about 75%) of those diagnosed with ADHD and an unknown proportion of children who do not meet the full criteria for a diagnosis of ADHD, but who have some ADHD symptoms, but not for any of the group labelled "Normals" who display little or no ADHD symptomatology. The circle represents those that actually do receive treatment with stimulants from clinicians. Included in the circle, are those who are inappropriately treated (over-prescription) and those who are appropriately treated (correct prescription). Outside the circle but to the right of the clinical response line, are those who would benefit, but who do not receive treatment (under-prescription). Clearly there is significant non-concordance between ADHD diagnostic criteria, prescription practices (the actual criteria applied by clinicians to make decisions about prescribing stimulants), and the degree of coverage of the population with a clinical need for prescription of stimulants.

It would therefore be beneficial to develop an assessment technique(s), which can more accurately predict a clinically beneficial response (all those to right of the vertical line). Most, but not all, children who meet ADHD diagnostic criteria respond to stimulant

medication. There is also anecdotal evidence from clinicians that some children with some symptoms of ADHD who do not meet the full criteria may also benefit.

One possible candidate for identification of responders and non-responders is the Continuous Performance Test (CPT), a neuropsychological test which has been available since it was first described by Rosvold et al [1954] and has been used clinically for children with ADHD since 1967 [Conners, Eisenberg & Barcai, 1967] and there are now several commercial versions available to clinicians (e.g. Gordon [1983], Conners [1992]). There are two major variants of the CPT: a simultaneous discrimination task where the subject responds by pressing a button to a target such as the letter 'X' appearing on screen; and a successive discrimination task, in which the subject responds to a target after the occurrence of a warning signal (e.g. the letter 'X' preceded by the letter 'A'). Variables measured include Errors of Commission (CPTC: i.e. button presses during non-target intervals), Errors of Omission (CPTO; i.e. number of targets omitted) and reaction time (RT; time in milliseconds between appearance of target on the screen and correct button press by the subject). A number of studies [Levy & Hobbes, 1981; Klorman et al 1991; and Levy & Hobbes, 1997]), indicate that the CPT is a good discriminator of children with ADHD from those without. The current practice is to compare a child's average score to a set of age-based norms, and if the child's scores are significantly (less than the 5th percentile) below the norm for age and gender, then this is an indicator that the child has attentional problems, which could in turn indicate ADHD.

The aim of the current study is to investigate the prediction of response to intervention with psycho-stimulant medication of children with ADHD using the CPT. The present hypothesis is that a non-linear classification by an MLP Neural Network will classify better than a linear Logistic Discriminant, in the prediction of response to treatment with stimulant medication. The prediction data set consists of the children's CPT scores, age and gender.

6.2 Method

Setting:

The sample for the study was retrospectively obtained from the clinical records for children seen and prescribed stimulant medication for ADHD by the Associate Professor Florence Levy, an experienced child psychiatrist specialising in the diagnosis and treatment of ADHD, at the Avoca Clinic, Prince of Wales Hospital for Children, Sydney, NSW, Australia.

Subjects:

	Whole Sample	Responders	Non-Responders
Male:Female	177:48	146:43	31:5
Average Age	8 years, 1 months	9 years	8 years, 3 months
Youngest Child	3 years, 11 months	3 years, 11 months	3 years, 11 months
Oldest Child	16 years, 3 months	16 years, 3 months	13 years, 7 months

Table 6.1 Sample Characteristics for the whole sample (n=225), Medication Responders (n=189) and Medication Non-Responders (n=36).

Four hundred and twenty one (421) clinical records of children initially being evaluated for ADHD and administered a Continuous Performance Test (CPT) by A/Prof Levy in the 5 years from 1994 to 1999 were reviewed. In one hundred and ninety six (196) of these children stimulant medication was not prescribed or a trial was commenced, but was abandoned too early to judge effectiveness. This left two hundred and twenty five (225) who completed a trial of stimulant medication. Of these one hundred and eighty nine (189) were judged on clinical follow-up to have responded positively to the medication and thirty six (36) were similarly judged to have not responded. Age and sex characteristics of the sample are presented in Table 6.1.

Procedure:

As part of A/Prof Levy's clinical protocol, pre-medication CPT scores were systematically obtained from children, by administration of a CPT test, where treatment with stimulant medication was under consideration. The CPT scores were stored on a database on the computer used to administer the CPT test. The CPT used is a computer-based assessment, which measures a child's speed of response, and errors to a task of recognising a target letter on the screen. One hundred and sixty (160) trials each lasting 1.5 seconds is presented to the child. Of the 160 trials twenty (20) present the target letter (in this case 'X'). The other 140 present other letters. The child responds to the letter 'X' appearing on the screen by pressing a button held in his/her hand. For trials in which the letter 'X' does not appear on the screen, the child is directed not to press the button. The

inter-trial interval is 1 second and the target trials are randomly interspersed amongst the non-target trials.

Available predictor variables for the study are listed in table 6.2 below:

Variable	Description
1 – 20	Twenty response times (in milliseconds)[1].
21	Errors of Omission (a count of targets missed altogether)
22	Error of Commission (a count of responses to non targets)
23	Blanks (a count of responses occurring during the inter-trial interval)
24	The child's age in months
25	The child's gender coded as 0= female, 1=male.

Table 6.2 The 25 Input (predictor) variables for Prediction of response to treatment with stimulant medication study.

[1] Note missed responses (an error of omission) were coded as 2000 milliseconds, since this value could not be obtained by responding (since it exceeds the 1500 millisecond length of the trial), but signifies a very slow response

The outcome variable was a binary categorical "responder/non responder" classification made by the treating child psychiatrist (A/Prof Florence Levy), who reviewed all clinical records of children considered for the study. This was assisted by the systematic practice by A/Prof Levy of reviewing progress at 4 to 6 weeks after initiating treatment and

attempting to ascertain success or failure of treatment and recording the response in the patient's clinical notes. A/Prof Levy was effectively 'blind' to the CPT scores at the time of making the responder/non responder classification. The CPT was administered at the first consultation (at least 4 to 6 weeks earlier) and it is automatically scored on age norms by the computer software used for administration. This produces a single score which places the child's overall performance (success and speed) in relation to his/her age peers. Clinically this is used to make the decision to prescribe a stimulant. Many children are not prescribed a stimulant, and therefore were not subjects in the present study. All the children included in the present study were prescribed Stimulant Medication because they all performed poorly on the single age adjusted CPT score and they met DSM-IV criteria for ADHD. A/Prof Levy did not have any access to the 23 raw CPT scores used as inputs for classification in this study, at any time during her clinical contact with a child or when making the outcome classification.

Classification Analyses

This data set was subjected to classification analyses with a Logistic Discriminant (LD) and eight MLP Neural Network with 2 to 9 hidden units. In addition, we trained an MLP without a hidden layer and without hidden units as a Logistic Discriminant (LD), for the classification problem.

All classification analyses were carried out using the Neural Network software, NEVPROP 4 (Goodman 1998), with Logistic Discrimination being implemented by a Neural Network that has no hidden unit layer and is trained using a Cross Entropy Error

Function (Goodman, 1998, Sarle, 1994). In such a case linear input units connect directly to a logistic output unit. All classifiers (LD and MLPs) were trained with QuickProp optimisation, Early Stopping with a 25% holdout, and Weight Decay (-0.01). These technical details are discussed in Chapter 4 and in Appendix 2.

As outlined in Chapter 4, the Akaike Information Criterion (AIC) was used to select one MLP model from amongst eight (MLP2 to MLP9) as the MLP model which will be directly compared to the LD model. AIC values are calculated using data from the training dataset. The MLP model with the lowest AIC value was selected as the MLP model to be used in comparison with the LD model.

Classification accuracy for both classifiers (LD model and selected MLP model) was measured using A_Z (Area Under the ROC Curve), calculated according to the method outlined by Harrel et al (1984). In order to obtain a measure of classification accuracy that can be generalised to the entire population from which the sample was drawn, bootstrapping (Efron & Tibshirani, 1993) was used to correct the estimate of A_Z. One hundred (100) bootstraps were used in each analysis.

6.2 Results

Model selection

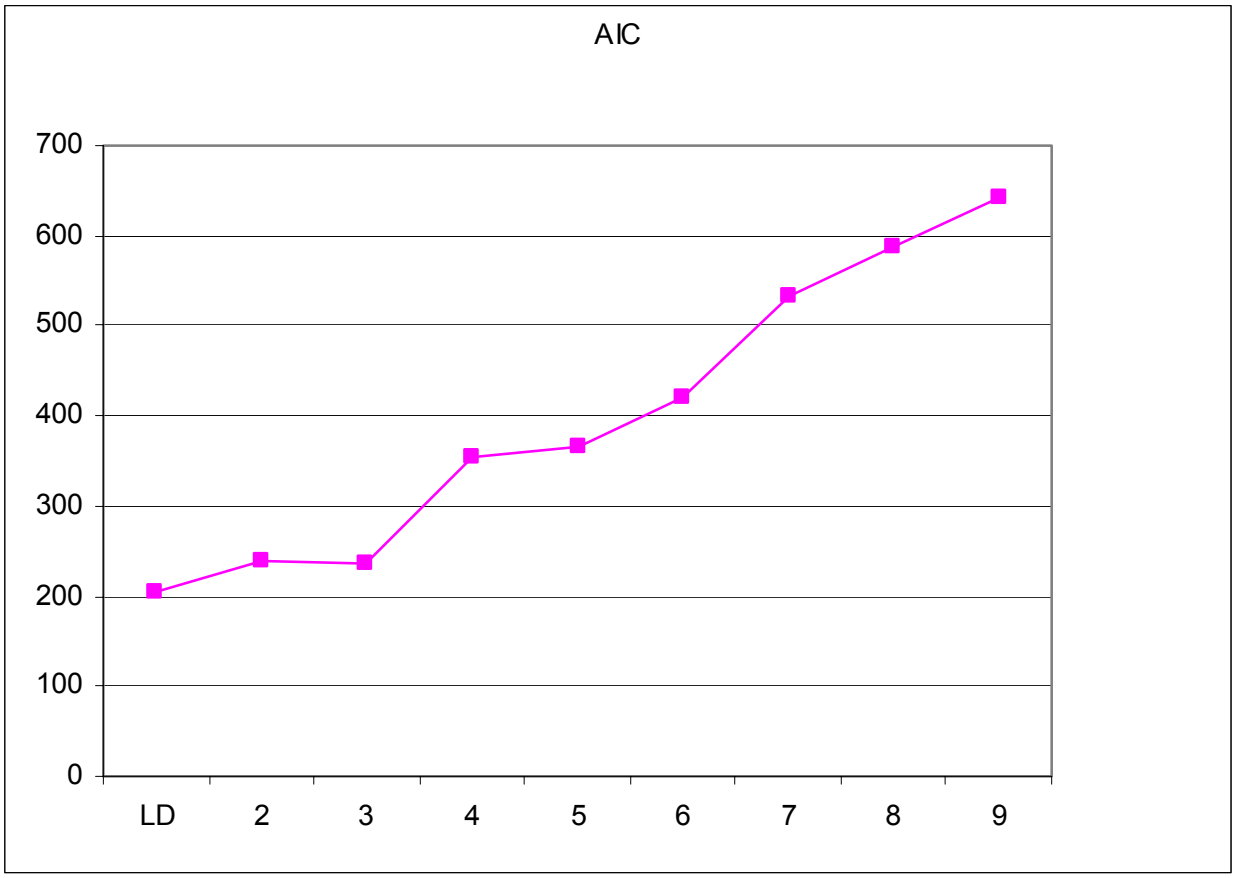

Figure 6.2 AIC values for LD and eight MLP (2 to 9 hidden units) models.

In Figure 6.2 above the MLP3 model is the MLP model with the lowest AIC value compared to all other MLP models. Thus the MLP3 model will be selected for comparison to the LD model.

LD Vs Selected MLP

Table 6.3 and Figure 6.3 below, present the classification accuracies of the Logistic Discriminant and the MLP Neural Networks (with 3 hidden units), in discriminating children who responded to treatment with stimulant medication from those who did not respond, using Age, Sex and pre-treatment scores on the CPT as predictor variables. Both the initial training dataset derived value and the bootstrap corrected estimate of the Area Under the ROC Curve are presented, along with the standard deviation of the bootstrap correction over 100 bootstrap samples.

Model	Training Dataset A_Z	Bootstrap Corrected A_Z	Std Dev of the Bootstrap A_Z (N =100)	Shrinkage A_Z
Logistic Discriminant	0.818	0.683	0.0328	.135
MLP 3	0.962	0.795	0.0200	.167

Table 6.3 Training Dataset and Bootstrap corrected A_Z (Areas under the ROC Curve) for Logistic Discriminant & MLP Neural Network (3 hidden units).

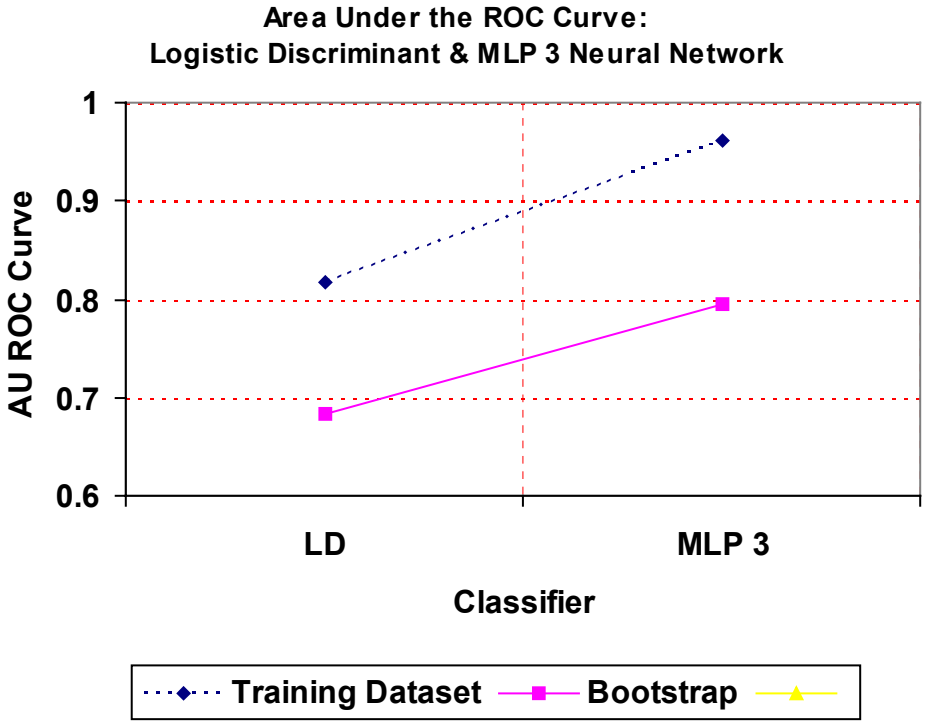

Figure 6.3 Training Dataset and Bootstrap corrected A_Z (Area Under the Receiver Operating Characteristics Curve) of Logistic Discriminant and MLP Neural Network Model (3 hidden units).

On both the initial estimate of accuracy and the bootstrap corrected estimate, the MLP 3 Neural Network discriminated responders from non-responders better than a Logistic Discriminant. The bootstrap corrected $\mathbf{A_Z}$ results indicate that the MLP with 3 hidden units, classifies better than the Logistic Discriminant. The difference (LD = .683 Vs MLP3 = .795) is statistically significant ($z = 2.08$, $p = .019$, $r_{pos} = .32$, $r_{neg} = .66$) using Hanley & McNeil's [1983] significance test, as outlined in Chapter 4, for comparing two A_Z.

The magnitude of shrinkage between the Training Dataset derived A_Z and the Bootstrap Corrected A_Z (0.167 A_Z units for MLP 3) observed in this study is very large. This is an indicator that all models are being over-fitted (see Appendix 2) and that the size of training dataset sample size is too small (see Study 4.8, Chapter 4). This is in spite of the implementation of measures (Early Stopping and Weight Decay) aimed to reduce over-fitting.

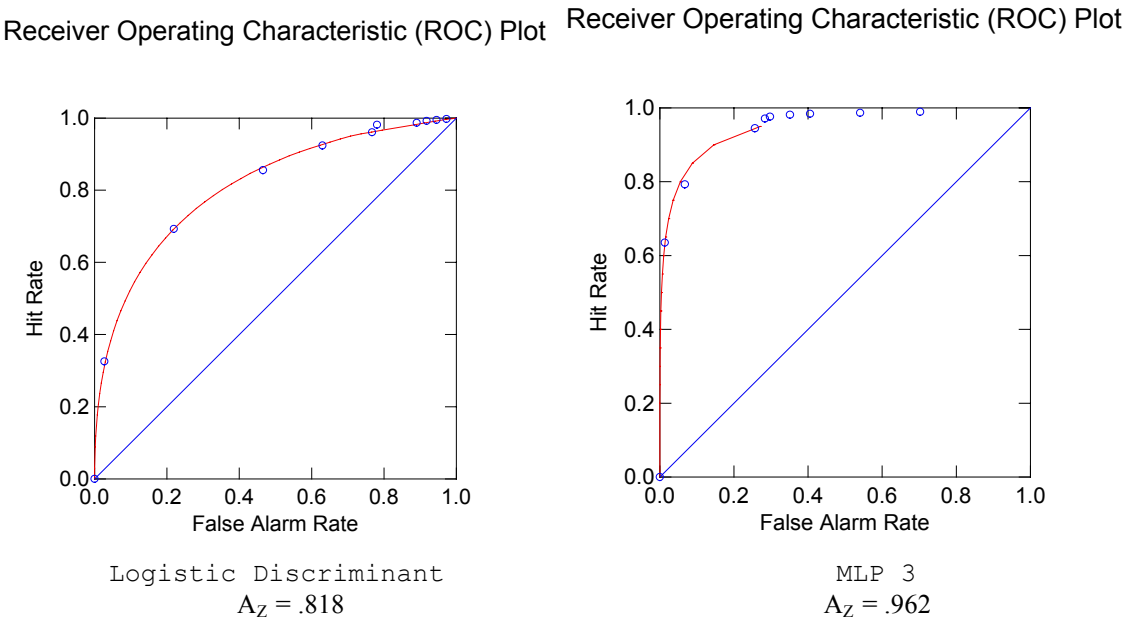

Figure 6.4 Training Dataset ROC plots for Logistic Discriminant and MLP 3

The Training data set ROC plots in Figure 6.4, demonstrate that the MLP 3 classified the Training dataset case better than the Logistic Discriminant.

6.3 Conclusions

Levy (Levy & Hobbes, 1981, Levy 1997) demonstrated that the Continuous Performance Test (CPT) is able to discriminate children with a diagnosis of ADHD from those without using linear discrimination techniques. The present study extends her result in two ways. Firstly it finds that the CPT is also able to discriminate, amongst children diagnosed with ADHD, children who clinically respond to treatment with stimulant medication from those who do not. Secondly both a linear technique (logistic regression) and a non-linear technique (MLP type Neural Network) were able to classify children with ADHD as responders or non-responders to treatment with stimulant medication, but the accuracy of classification by the MLP was better than that of the Logistic Regression using this same set of predictors, criterion and sample of treated children. This indicates that the Bayesian classification boundary between the class dependent distributions of responders and non-responders, in a data space defined by 25 variables (age, sex and pre-treatment scores on the CPT) is non-linear.

This finding supports the main hypothesis of this thesis that some clinical decision making problems in Psychiatry can be more advantageously solved by MLP type Neural Networks than by traditionally used linear statistical techniques, because in these cases the classification problem has an inherent non-linear Bayesian decision boundary.

Limitations

From the point of view of developing a clinical application, the findings are indicative only. They have internal validity in that the comparison is valid for this sample and for the criteria used, but external validity is compromised by four major limitations.

First, the total sample size (n=225) is small, but in particular the number amongst the sample who where non-responders (n=36) is very small. The degree of shrinkage observed in the study is very large. This raises an issue about generalisation to a larger population of non-responders. A much larger training dataset sample size is required to generate a better classification solution to this problem.

A second limitation is the use of clinical judgement by a single clinician as the criteria for a clinical response. The study was retrospective and we only had access to clinical notes to for data collection to construct our response to treatment variable. Our use this criteria however raises an issue about external validity, because the judgements of other clinicians might not agree with those of our single clinician. The study was designed primarily to identify if response to treatment with stimulant medication in children with ADHD was a potential candidate application for a Neural Network approach. The findings suggest that it is. The next step would be to design a prospective study and use more objective outcome measures such as parent and teacher rating scales or structured observations of behaviour and/or symptoms.

A third limitation may be that the classification problem could be "better posed" by the addition of more predictor variables. In addition to CPT scores, age and sex, other candidate predictors such as: DSM-IV symptom data; DSM-IV ADHD subtypes; parental ratings of ADHD symptom ratings; teacher ratings of ADHD symptoms; or cognitive test scores, could be considered. The addition of some of the above may result in improved prediction of clinical response to stimulants. However, in order to conduct such investigations, much larger datasets than that employed in the current study would be required.

Finally in Figure 8.1 we indicated that there may exist a group of children, who do not qualify for a diagnosis of ADHD but who have some symptoms of ADHD (ADHD – NOS). The subjects in this study all had a DSM-IV diagnosis of ADHD. Therefore the finding cannot be directly generalised to this group. Future investigations of the capacity of the CPT and MLPs to predict the response to stimulant medication in this group are suggested by the findings of the present study.

The main value of this study is to suggest that further research on the problem of prediction of response to treatment with stimulant medication in children with ADHD, using a neural network approach is warranted. Because we did not have an independent test set, we are not able to assess (quantify) how well the MLP model will classify future cases.