Florio, T., Einfeld, S., Tonge, B., & Brereton, A. (2009). Providing an Independent Second Opinion for the Diagnosis of Autism Using Artificial Intelligence over the Internet. *Counselling, Psychotherapy, and Health*, 5(1), The Use of Technology in Mental Health Special Issue, 232-248.

Counselling, Psychotherapy, and Health
www.cphjournal.com
cphJournal.com

Providing an Independent Second Opinion for the Diagnosis of Autism Using Artificial Intelligence over the Internet

Tony Florio [1, 2]
Steward Einfeld [2]
Bruce Tonge [3]
Avril Brereton [3]

Contact:
Tony Floriao
tony@florio.com.au

1 The Rozelle Hospital, NSW
2 School of Psychiatry, University of NSW
3 Developmental Psychiatry Monash Medical Centre, Melbourne, VIC

Abstract

In many fields, including the diagnosis of Autism, there are barriers which greatly reduce the likelihood of clinicians obtaining an independent second opinion. Barriers include cost and the availability of other specialist clinicians. Nevertheless, having regular independent second opinions for the diagnosis of Autism is professionally indicated. The use of new technology as a means of providing an independent second diagnostic impression for Autism is reviewed in this article. The use of technology may provide independent clinical assessment thus adding to diagnostic accuracy. Challenges regarding the use of technology to provide independent second opinions include the possibility of individuals and relatives engaging in self diagnostic activities.

Independent Second Opinions

It is good clinical practice for clinicians to get an independent second opinion when they are not certain about a diagnosis. Even if a clinician is relatively certain of their diagnosis an independent second opinion in may give them cause to think again. Ideally the clinician obtains the independent second opinion from another clinician who is a recognised expert in the area of diagnosis in question. Also ideally the second opinion clinician is blind to the diagnosis, thoughts and actions of the first opinion clinician and so is therefore unbiased, by these, in making their diagnosis.

Given the benefits of clinicians obtaining an independent second opinion, why then does this practice not occur more often. In many fields (the diagnosis of Autism is one of them) there are barriers which greatly reduce the likelihood of clinicians obtaining an independent second opinion. In the case of the diagnosis of Autism, these barriers are:

- A second opinion from a second clinician necessarily involves, the client (and in most cases their family) attending the practice of the second clinician on a second occasion. This can be both inconvenient and costly for the family
- Two opinions will generally cost twice as much as one opinion. Someone, whether it be the client, their family, a government or some combination, will in the end pay this bill and this naturally raises a question in the mind of this payer as to the amount of increased benefit brought about by a double cost.
- An “independent” second opinion by a second clinician in the case of a diagnosis like Autism may be not be practical to obtain. There is foreseeable risk that family members who attend the second clinician’s assessment will

"leak" or at least hint the first clinicians opinions or diagnosis to the second clinician.

- Finally, clinical experts in any area (and Autism Diagnosis is no exception) are thin on the ground. Those that exist are usually busy providing first opinions to their clients and would not realistically have the time to provide numerous second opinions to their colleagues. If obtaining an independent second opinion from a second expert clinician becomes a common practice, then there will need to be an increase in the number of practicing expert clinician which is directly proportionate to the increase in this practice.

Despite these barriers, obtaining an independent second opinion for difficult diagnoses remains a desirable goal. In some clinical areas, such as Anatomical Pathology second opinions are routine sought. It is common practice in the screening of Pap Smears for cervical cancer for all cases identified as positive and a randomly selected proportion of cases identified as negative by a first clinician to be re-screened by a second clinician.

This article describes a method for providing an independent second opinion in the case Autism Diagnosis.

Diagnosis of Autistic Disorder

Autistic Disorder is a defined developmental disorder (DSM-IV, APA 1994), which is characterised by:

- Onset before the age of 2 years, often become evident by then.
- poor social interaction compared to aged peers
- language which both delayed in development and disordered in its form.
- a range of odd behaviours

- a profile of uneven mental abilities (marked differences in strengths and weaknesses)

It is a serious disorder, it is lifelong and it is disabling. Persons with Autistic disorder need special education during school aged years and will generally need support as adults. Importantly early diagnosis is crucial, because symptoms and subsequent disability can be ameliorated by early intervention. The earlier intervention can be provided the better the long term outcome. The provision of intensive early intervention to children with Autistic Disorder is expensive and a diagnosis provided by an appropriate clinical expert (child psychiatrist, developmental pediatrician or psychologist) is often an essential pre-requisite.

Autistic Disorder has a prevalence of 15 – 50 per 10,000, in very young children, presents similarly to a range of other Autistic Spectrum Disorders (prevalence 1 in 100) and to some other Developmental Disorders (prevalence 2 – 3 per 100). Diagnosis is not straight forward at very early ages and expert clinicians often experience uncertainty.

Developmental Behaviour Checklist (DBC)

A 96 item *parent completed* checklist, developed by Einfeld & Tonge (1995), with funding from National Health & Medical Research Council. The DBC assesses behavioural and emotional problems in children and Adolescents with a Developmental Disorder (age 2 to 18 years). Parents or carers are asked to rate a child on 96 behavioural descriptors using a three point scale (0,1,2).

The DBC has good reliability and validity, and it has been used in numerous studies published in peer reviewed scientific journals.

The DBC is primarily used to describe and quantify behavioral and emotional problems in children and adolescents with a developmental disorder. A total score (interpreted via a set of norms) is used to quantify the overall severity of the child's behavioural and emotional problems a set of six subscales (derived by factor analysis) is used to profile and quantify (via norms) specific areas of concern for the child.

More recently, the ability of information captured by the DBC to aid in diagnostic decision making for Autistic Disorder, particularly in very young children, has been explored.

Artificial Neural Networks

Artificial Neural Networks were initially developed by researchers attempting to simulate the functioning of neurons in the Central Nervous System (CNS) of animals. The CNS is effectively a network of neurons. One of the questions researchers and theorists have attempted to answer is where in the CNS is learning stored when an animal learns a new skill or behaviour. Hebb's (1949) theory was that the synapse (the point at which one neuron connects to another neuron) is the key location, within a biological neural network where new learning is stored. More specifically he proposed that changes to the strength (propensity to transmit a signal from one neuron to another) of the connections between neurons in a network underlie learning by animals that have a CNS.

Motivated to test Hebb's theory, experimenters in the 1960s, 1970s simulated neurons (at first mechanically, but later as software emulations on a computer) and interconnected these into networks with topologies (connection schemas) similar to those which have been observed in vivo. It was first demonstrated that the behaviour of each network as a whole (or more specifically the relationship between the inputs and the outputs to a network) was always consistent in that the same input always produced the same output, but different inputs could produce different outputs. That is each network had its own consistent set of input-output pairs. Furthermore it was demonstrated that by altering the strength of synaptic connections in a particular network, its behaviour could be altered, so that it exhibited a different set of input-output pairings. In learning, such as is demonstrated by classical conditioning or by operant conditioning, animals (including humans), learn to produce a particular behavioural response (an output of the CNS) in response to a particular stimulus (an input to the CNS). The basic problem of how exactly the CNS does this has (and still does) puzzle researchers and theorists.

Working more in a statistical framework, Werbos (1974) developed a technique, known as back-error propagation, which can be used to adjust the strength of synaptic connections between artificial neurons organized into a particular network topology known as a Multi-Layer Perceptron (MLP), so as to make the MLP network learn from experience. The basic process is that the MLP is presented an input and in response the MLP generates an output. The output is the direct result of the input being processed by the MLP which transmits information from neuron to neuron using a defined set of weights. Each weight represents the synaptic connection strength between a specific pair of neurons in the MLP. As well an input the MLP is presented with a desired output (that is how the MLP should have responded to this particular input). The back-error propagation process begins by quantifying the amount of error (or difference) between the output the MLP actually produced and the output the MLP should have produced. If the difference between actual and desired is large then the quantitative size of error is large. If the difference between actual and desired is small then the quantitative size of error is also small. Working backwards through the network, the back-error propagation algorithm then adjusts each weight in the MLP by an amount proportionate to the size of the error (a large error leads to large adjustments and a small error leads to small adjustments) so that next time the same input is presented the size of the error between the actual output and the desired output is reduced. If the same input-output pair is presented and back-error propagated many times, the network eventually learns to produce the desired output in response to the specified input. If there is a set (several or many) of input-output pairs and they are individually presented and back-error propagated to an MLP network many times, then the MLP will eventually learn the whole set an always produce the correct output response for each specific input.

The upshot of all this is that it has been empirically demonstrated that MLP neural networks using a back-error propagation algorithm for weights adjustments can learn any finite set input-output relationships. White (1989) later provided a mathematical proof of this proposition. Incidentally these findings have been statistically interesting and have led directly to the development of many statistical applications based upon MLP neural networks (e.g. the Autism Diagnosis application discussed later in this article), but they proved to be a dead end in testing Hebb’s theory because the concept of back-error propagation is biologically implausible. In order for a mechanism similar to back-error propagation to exist in the CNS there would need to be two way transmission of information between neurons. Our current knowledge suggests that biological neurons are able to transmit information only in one direction.

The importance of MLPs and back-error propagation in the current context is that it allows for the possibility of creating an artificially intelligent system, which can be trained to make a diagnosis of Autistic Disorder (the output) on the basis of a set behavioural symptoms such as DBC items (the inputs), by using a set of cases organized as a set of input-output pairs.

For a more detailed exposition of MLP neural networks, back-error propagation and their uses in diagnostic decision making see Florio et al. (1994), Cross et al. (1995), Florio et al. (1997) or Florio (2004).

Diagnosis of Autistic Disorder using a Neural Network

From the Monash Medical Centre Autism Clinic in Melbourne and the Australian Child Development Study, we obtained 638 cases, 50% with Autistic Disorder, 50% with a Developmental Disorder but not Autistic Disorder. Expert clinicians (Child Psychiatrist or Clinical Psychologist) who were experienced in Autism made all the ‘gold standard’ diagnosis used to train the Neural Net.

Secondly we obtained another cross-validation dataset of 100 cases (62% Autistic Disorder, 38% Developmental Disorder but not Autistic Disorder) collected in Sydney, at 3 separate child development clinics (Leichhardt, Tumbatin and Kogarah clinics).

For both datasets the variables were 96 parent\carer completed DBC items, Age, Sex, IQ level (coded as severely disabled, moderately disabled, mildly disabled, or normal range) and Autistic Disorder Diagnosis (coded yes, no according to an expert clinician diagnosis using DSM-IV criteria)

The aim of the study was to compare a MLP type Neural Network to a Logistic Regression (Hosmer & Lemeshow, 1989), to see if the MLP Neural Net was any better as a diagnostic classifier.

For both the Neural Network and the Logistic Regression the first dataset (N=658) was used to derive diagnostic classifiers. Area under the ROC curve (Swets, 1988) was used as the comparative measure. In both cases the Area under the ROC curve was adjusted (attenuated for optimism) by a 100 x bootstrapping procedure (Efron & Tibsharani, 1993). The adjusted ROC curve value gives a good approximation to how each diagnostic classifier will perform on future cases.

The second dataset (N=100) was used to comparatively evaluate the performance of both classifiers (Neural Network and Logistic Regression) on set cases not previously used to derive either classifier (cross-validation).

Results - Neural Net Vs Logistic Regression

For the Melbourne dataset (N = 638)

	Neural Net	**Logistic Regression**
% Correct	**92%**	**83%**
ROC Curve	**.98**	**.91**
Adjusted ROC	**.93**	**.88***

* Difference significant at p = .001, using the Wilcoxon Test

Conclusion

The Neural Net is a better classifier than Logistic Regression for this particular diagnostic problem

Independent Multi-site Cross-validation

In order to assess how well the Neural Net would perform in 'real life' at different clinics, we tested it on further 100 cases (62% with Autistic Disorder) seen at 3 Sydney child development clinics (Grosvenor, Tumbatin and Kogarah). Expert clinicians (Developmental Pediatrician or Clinical Psychologist) who were experienced in Autism made all the 'gold standard' diagnosis against which the Neural Net was compared.

Overall accuracy	**80%**
Sensitivity	**92%**
Specificity	**70%**
ROC Curve	**.88***

*** Note** in this analysis, because we are using a cross-validation dataset, we do not need to adjust the Area under the ROC Curve value by bootstrapping.

Conclusion

In the real world, the DBC-Neural Net diagnosis of Autistic Disorder is 80% as accurate as that provided by very experienced clinicians.

Comparison to Other Checklists and Techniques

Clinicians already have a range of diagnostic tools in this area

Method\Practice	Validity	Cross-Validity
DSM-IV	**Not Applicable**	**85% (agreement between Clinicians)**
CARS	**82%**	**?**
GARS	**90%**	**?**
ADOS	**95%**	**?**
ADI-R	**92%**	**?**
DBC-ASA	**78%**	**?**
DBC-NN	**92%**	**80 %**

Conclusion

The DBC-NN compares favourably with other commonly used tools and practices in terms of validity. None of the other tools has used a cross-validation test, which gives the most accurate index of a diagnostic classifier's ability to classify. However agreement between clinicians on a diagnosis of Autistic Disorder was found to be 85% in the DSM-IV field trial studies. Since a clinical diagnosis using DSM-IV criteria is the gold standard for diagnosis of Autistic Disorder, then 85% is the maximum possible cross-validity agreement any other method could have. The DBC-NN at 80% is close to this theoretical ceiling and therefore performing relatively well.

Summary DBC-NN

- ☐ Parent completes a checklist.
- ☐ The DBC-Neural Network makes an Autistic Disorder Diagnostic decision which is at least 80% as good as that of any very experienced expert clinician.
- ☐ It is almost as good as having a second independent clinician, as on average clinicians will only agree with each other on a diagnosis of Autistic Disorder 85% of the time, when blind to the diagnoses of other clinicians (DSM-IV field trial).
- ☐ Any clinician can use the DBC-NN as a blind independent 2nd opinion.
- ☐ Unlike most other methods, the DBC-NN is totally independent of the clinician, as it is based on independent parent observations.

Internet Diagnosis

We currently sell a software package to clinicians, but this is problematic because:

- we need to physically distribute disks to users
- provide users support with installation issues
- can only target one platform (Windows)

- Can only upgrade users every few years due to effort involved

A natural next step is to use the Internet to give clinicians access. In this mode:

- The parent or carer completes the DBC checklist online
- Clicks a submit button
- A comprehensive report, including an Autistic Disorder diagnostic decision is either sent back as a web page or emailed to the clinician.

We see other advantages to using the net. We can update the application as required without a distribution cost. We can accumulate datasets which:

- Tell us how our diagnostic application is being used, how often, by whom and with which clients.
- Can be used to update our normative datasets

We can also continuously refine the diagnostic accuracy of the application by retraining the neural network on progressively larger datasets, if clinicians provide us their first opinion diagnosis. This akin, to how a clinician gains experience and refines their diagnostic skills in the course of a clinical career. The main difference is that in 30 year career a clinician may see at most a few thousand cases. The neural network could see tens of thousands in a year. From this ‘experience’ base and trained by the first opinion diagnoses of hundreds of expert clinicians from around the world, a popularly used neural network would become the best diagnostician for this particular diagnosis over time. This has already happened in the case of PAPNET (Kok & Boon, 1995) a neural network which diagnoses abnormalities in PAP smears, which has progressed from being used as second opinion system to a first opinion in some health services (Cenci et al 2000).

Challenges

- Avoiding 'self'-diagnosis by parents
- Avoiding inappropriate use by clinicians, in particular screening for Autism, which will yield a large proportion of 'False Positives'
- Developing an e-commerce model that:
  - works for us,
  - works for clinicians and
  - works for clients
- Overcoming 'change in practice' inertia amongst clinicians.
- Issues about who 'owns' the data, do we save it, do we store it for clinicians, what ownership rights do clients (parents) have, how do we get consent for ongoing development ?

Future Developments

If we store data and get selected clinicians to send us their diagnosis, we have a continuing supply of new cases, with DBC data and high quality diagnoses. We can use these to continuously, incrementally improve diagnostic accuracy, by periodically retraining the neural net on ever larger datasets

Potentially over time the neural net absorbs the collective diagnostic expertise of a large number of clinicians from around the world. The DBC-NN can then potentially move from being a 2nd opinion system to becoming 1st opinion system (this is happening now with PAPNET, see Cenci et al [2000]).

AI - Internet → Synergy

Artificial Intelligence (AI) and the Internet are both recently emerged computer-based technologies. As such there is a natural opportunity for synergy.

Neural Networks, are computer-based AIs which can learn from experience. If we give a neural network a large number of real-world experiences and at the same time give them feedback as to how they should best respond to each of these experiences, then the neural network can learn to solve real-world problems. What they learn is how to best respond to different real world generated stimuli. For example in this article we describe how a neural network was trained to respond (with 80% accuracy compared to an expert clinician) to a set of parent\carer ratings of a child's behaviour with a diagnostic decision as whether the child has a diagnosis of Autistic Disorder or not. Numerous other examples of diagnostic decision-making by neural networks can be found in Florio et al (1994), Florio et al (1997), Florio (2004).

The Internet flexibly connects computer systems from millions of locations around the world. So the Internet potentially gives an AI access to a large number of real-world experiences. In the virtual space of the Internet a single AI is able to make numerous and distant real world connections

Therefore an Internet based AI which is able to receive real-world exposure and feedback, in a particular domain is likely to continuously learn to better respond to that domain of experiences it is exposed to, in much the same way as a clinician, professional, tradesperson or other 'expert' learns on the job with the help of supervisors and colleagues.

In our case, Autism Diagnosis, we want to develop a Neural Network AI, which is able to continuously improve, as it practices (and learns), seeing thousands of patients in hundreds of clinics and absorbing the diagnostic wisdom of hundreds of experienced clinician 'colleagues'.

Information & Links

- We are attempting to develop the site now, Details will be announced on http://florio.com.au

- The Developmental Behaviour Checklist homepage is: http://www.med.monash.edu.au/psychmed/units/devpsych/dbc.html

- Annotated bibliography of research and publications using the DBC: http://www.med.monash.edu.au/psychmed/units/devpsych/bibliography.pdf

- Similar sites up & running now are:
  - http://www.prostatecalculator.org/
  - http://www.aseba.org/PRODUCTS/weblink.html (CBCL)

- A full downloadable copy of Tony Florios's PhD thesis containing a full exposition of Neural Networks and the development of the DBC-NN is available at: http://florio.com.au

References

American Psychiatric Association (1994) Diagnostic and Statistical Manual of Mental Disorders (4th ed.) Washington DC, APA

Brereton AV, Tonge BJ, Mackinnon AJ & Einfeld SL (2002). Screening Young People for Autism With the Developmental Behaviour Checklist. Journal of the American Academy Child & Adolescent Psychiatry; 41(11): 1369-1375

Cenci M. Nagar C. Vecchione A. PAPNET-assisted primary screening of conventional cervical smears Anticancer Research. *20(5C):3887-9, 2000 Sep-Oct.*

Cross S.S., Harrison R.F., & Kennedy R.L. (1995) Introduction to Neural Networks, Lancet, 346(8982):1075-1079, Oct 21

Efron B. & Tibshirani R.J. (1993) An Introduction to the Bootstrap, Chapman & Hall, NY.

Einfeld S.L. & Tonge B.J. (1995) The Developmental Behaviour Checklist-The development and validation of an instrument to assess behavioural and emotional disturbance in children and adolescents with mental retardation, Journal of Autism and Developmental Disabilities, 25(2), 81-104

Florio T., Einfeld S.L. & Levy F. (1994), Neural Networks and Psychiatry: Candidate Application in Clinical Decision Making Australian & New Zealand Journal of Psychiatry, 28(4):651-666.

Florio TM. Parker G. Austin MP. Hickie I. Mitchell P. Wilhelm K**.** Neural network subtyping of depression. Australian & New Zealand Journal of Psychiatry. *32(5):687-94, 1998 Oct.*

Florio T (2004) The use of Neural Networks for Clinical Decision Making in Psychiatry. PhD thesis, School of Psychiatry, University of NSW. Full copy available as a set of PDF files at http://florio.com.au

Hebb DO. The Organisation of Behavior: A Neuropsychological Theory. New York: John Wiley, 1949.

Hosmer DW & Lemeshow S (1989) Applied Logistic Regression, Wiley, NY, 1989

Kok MR & Boon ME (1996) Consequences of neural network technology for cervical screening: increase in diagnostic consistency and positive scores. Cancer. *78(1):112-7, 1996 Jul 1.*

Swets J.A. (1988) Measuring the Accuracy of Diagnostic Systems. Science, 1988, June 3: 1285-1293.

Werbos P. Beyond Regression: new tools for prediction and analysis in the behavioural sciences. PhD Thesis Harvard University, 1974

White H. (1989) Learning in Artificial Neural Networks: A Statistical Perspective. Neural Computation ,1989, 1, 425-464