European Journal of
Nuclear
Medicine

Original article

Artificial neural networks that use single-photon emission tomography to identify patients with probable Alzheimer's disease

M.R.W. Dawson[1], A. Dobbs[1], H.R. Hooper[2], A.J.B. McEwan[3], J. Triscott[4], J. Cooney[5]

[1] Department of Psychology, University of Alberta, Edmonton, Alberta, Canada
[2] Department of Nuclear Medicine, Cross Cancer Institute, Edmonton, Alberta, Canada
[3] Department of Radiology and Diagnostic Imaging, University of Alberta, Edmonton, Alberta, Canada
[4] Department of Family Medicine and Division of Geriatric Medicine, University of Alberta, Edmonton, Alberta, Canada
[5] Department of Psychiatry, University of Alberta, Edmonton, Alberta, Canada

Received 3 February and in revised form 7 July 1994

Abstract. Single-photon emission tomographic (SPET) images using technetium-99m labelled hexamethylpropylene amine oxime were obtained from 97 patients diagnosed as having Alzheimer's disease, as well as from a comparison group of 64 normal subjects. Multiple linear regression was used to predict subject type (Alzheimer's vs comparison) using scintillation counts from 14 different brain regions as predictors. These results were disappointing: the regression equation accounted for only 33.5% of the variance between subjects. However, the same data were also used to train parallel distributed processing (PDP) networks of different sizes to classify subjects. In general, the PDP networks accounted for substantially more (up to 95%) of the variance in the data, and in many instances were able to distinguish perfectly between the two subjects. These results suggest two conclusions. First, SPET images do provide sufficient information to distinguish patients with Alzheimer's disease from a normal comparison group. Second, to access this diagnostic information, it appears that one must take advantage of the ability of PDP networks to detect higher-order nonlinear relationships among the predictor variables.

Key words: Artificial neural networks – Single-photon emission tomography – Alzheimer's disease

Eur J Nucl Med (1994) 21:1303–1311

Introduction

Investigations of the applicability of single-photon emission tomography (SPET) to the understanding and the diagnosis of patients with probable Alzheimer's disease (pAD) have led to reports of encouraging results by a number of authors [e.g. 1–8]. First, using a variety of tracers, many researchers have found that perfusion in the temporal and parietal lobes is characteristically lower in pAD patients than in a variety of comparison groups. Second, some authors have shown that quantitative measures of perfusion from SPET can demonstrate strong positive correlations with a variety of behavioral measures of disease severity [e.g. 5, Fig. 4]. Third, the abnormalities revealed by SPET images of pAD patients tend to be quite consistent with those revealed by positron emission tomography (PET). This last finding is important in that SPET imaging techniques are less expensive and as a result are more widely available than PET techniques. Form a clinical perspective, SPET appears to have significant potential for identifying patients with Alzheimer's disease.

However, the results cited above primarily represent average differences between groups of patients, or between patient groups and control groups. When the emphasis shifts to using SPET to determine whether a particular individual has Alzheimer's disease, the results are not as compelling. In a recent literature review, Albert and Lafleche [9] noted that the accuracy of diagnosis on the basis of SPET measures is highly variable, and is strongly dependent on disease severity. For example, accuracy in discriminating severely impaired patients from controls is typically very high, ranging from 85% to 100%, In contrast, discriminating mildly impaired patients from controls has resulted in accuracy rates as low as 25%. Albert and Lafleche concluded that PET and even computed tomography (CT) are generally more accurate than SPET.

This paper describes a new quantitative technique for using SPET measurements to diagnose pAD. The quantitative technique that we utilize is a particular *artifical*

Correspondence to: M.R.W. Dawson, Biological Computation Project, Department of Psychology, University of Alberta, Edmonton, Alberta, Canada T6G 2E9

neural network (ANN), namely the network of value units originally described by Dawson and Schopflocher [10]. With any quantitative technique for medical diagnosis, two separate issues must be considered. First, it must be shown that the technique is capable of performing a differential diagnosis for patient/non-patient samples. Second, *after* it has been established that the technique can distinguish patients from non-patients, it must then be shown that this diagnostic capability generalizes to new patients. This paper is primarily concerned with establishing the first property for ANNs, namely the ability to correctly discriminate pAD patients from a comparison group in a large sample of subjects.

This paper proceeds as follows: First, we argue that ANNs may be required for pAD diagnosis using SPET, because this diagnosis may be a "hard" (i.e., not a linearly separable) pattern recognition task [see II]. Second, we briefly describe the basic properties of an ANN, and show how in principle its pattern recognition abilities are more powerful than those of other quantitative techniques, such as multiple regression. Third, we present the results of a study that compared the diagnostic abilities of ANNs to those of multiple regression. The results of this study show that ANNs are significantly better quantitative tools for pAD diagnosis. Indeed, we show that in some cases an ANN can use SPET data do *perfectly* discriminate pAD patients from healthy comparison subjects, even when a large, heterogeneous sample is used (97 pAD patients and comparison subjects).

Why use ANNs for pAD diagnosis?

SPET diagnosis of pAD as a "hard" pattern recognition problem. Computer vision researchers are often interested in developing programs to recognize or classify input patterns [e.g. 11]. Clearly, the use of SPET images to diagnose specific diseases can be viewed from this pattern recognition framework. For example, in the research reported below, each patient is described as a pattern of 14 different features, where each feature is the number of scintillation counts detected by SPET measurement of a specific brain region. The goal of a pattern recognition system would be to classify a particular patient (i.e., to label them as normal or as pAD) on the basis of these input features.

Importantly, not all pattern recognition tasks have the same level of difficulty. Some pattern recognition tasks are very easy, because a patient can be classified as being a member of particular group by noting the presence or absence of a definite feature. For example, let us imaginge that using SPET to diagnose pAD was this kind of task. This would mean that the presence of one feature (say, reduced temporal lobe perfusion) would indicate that the subject had probable Alzheimer's disease, and the absence of this feature would indicate this was not the case. In general, an easy pattern recognition task like this is *linearly separable*. For instance, in our example each patient could be represented as a single point in a 14-dimensional space, where their coordinate in each dimension of this space was the value of a specific SPET feature. In a linearly separable problem, a single straight cut (i.e. a hyperplane) could be made through this space that would separate all the pAD patients from all the non-pAD patients. In our imaginary example, this cut would separate all the patients with low temporal lobe perfusion from those with normal temporal lobe perfusion.

Previous studies that have attempted to diagnose pAD using SPET appear to have implicitly adopted the assumption that this pattern recognition problem is "easy", in the technical sense that it is linearly separable. For instance, many researchers have attempted to discover a single positive feature (e.g. some quantitative index of temporal or parietal perfusion capable of discriminating pAD patients from comparison groups [e.g. 1, 4, 5]. Furthermore, most studies have used very small numbers of patients, and, in some cases, have also used prescreening methods that have ensured that these small groups are quite homogeneous [e.g. 5]. In general, pattern recognition systems trained on a small number of instances will only correctly classify new patterns if they are dealing with a linearly separable class [e.g. 12, 13].

Unfortunately, it is not at all clear that a SPET basis for the diagnosis of pAD is "easy" or linearly separable. This is because there is a good deal of neurological [e.g. 14] and behavioural [15, 16] variability within the population of pAD patients. As a result, there may *not* be a definite quantitative feature that leads to a unique and correct diagnosis. Perhaps SPET's poorer diagnostic utility relative to other imaging techniques [9] may simply reflect that researchers have mistakenly treated the identification of pAD patients as a linearly separable problem. If this is the case, then more accurate diagnosis will require that complex interactions among many different input features be considered. Furthermore, more than one of these interactions may be indicative of pAD. In short, the diagnosis of pAD using SPET may be a "hard" or *linearly non-separable problem*, and as a result may require novel quantitative approaches in order to be solved.

In general, the solution to a linearly non-separable problem requires that many different cuts be made through a pattern space in order to separate members of one class from the members of another. Because of this, standard statistical techniques (e.g. multiple regression) which are capable of discovering definite features that solve linearly separable pattern recognition problems are very poor at discovering solutions to linearly non-separable problems (see below). In contrast, a modern ANN is admirably suited to solving linearly non-separable problems [e.g. 17] because it can use internal feature detectors to discover multiple non-linear interactions among a set of input variables, which in turn permit a complex partitioning of the pattern space.

Introduction to ANNs. An ANN is a computer simulation of a "brain-like" system of interconnected processing

units (Fig. 1). In general, an ANN can be viewed as a multiple-layer system for generating a desired response to an input stimulus. The stimulus is provided by the environment, and is encoded as a pattern of activity in a set of *input units*. The response of the system is represented as a pattern of activity in the network's *output units*. Intervening layers of processors, called *hidden units*, detect higher-order regularities in the input stimulus that allow the network to make a correct or appropriate response. It is these hidden units that result in the multiple partitioning of a pattern space, permitting the solution of linearly non-separable problems.

Processing units in an ANN are typically viewed as being analogous to neurons, and are presumed to operate in parallel. The behaviour of a single processing unit in this system can be characterized as follows: First, the unit computes the total signal being sent to it by other processors in the network. Second, the unit adopts a particular level of internal activation on the (non-linear) basis of this computed signal. Third, the unit sends a signal based on this internal activity to other processors in the network.

The signal that one processor sends to another is transmitted through a weighted connection, which is typically described as being analogous to a synapse. The connection itself is merely a communication channel. The weight associated with this connection defines the nature and strength of the connection. For example, inhibitory connections are defined with negative weights, and excitatory connections are defined with positive weights. The strength of the connection is defined by the size (i.e. absolute value) of the weight. The pattern of connectivity in an ANN (i.e. the networks's entire set of connection weights) defines the causal relations between the network's processors and is therefore analogous to a program in a conventional computer [e.g. 18].

However, in contrast to a conventional computer, the ANN is not given a step by step procedure to perform some desired task. Instead, the network is *taught* to do the task. For example, consider a popular supervised learning procedure called the *generalized delta rule* [e.g. 19]. To train a system with this rule, one starts with a network (of a prespecified number of processing units) that has small, randomly assigned connection weights. The network is then "developed" by presenting it a set of training patterns, each of which is associated with a known correct response. To train a network on one of these patterns, the pattern is presented to the network's input units, and the network generates a response to this stimulus using its existing connection weights. An error value for each output unit is generated by comparing the actual output to the desired output. This error value is then fed backwards through the network, and is used to modify connection weights in such a way that the next time this pattern is presented to the network, the network's output errors will be smaller. By repeating this procedure a large number of times for each pattern in the training set, the network's response errors for each pattern can be reduced to near zero. At the end of this procedure, the network will have a very specific pattern of connectivity (in comparison to its random start) and will have learned to perform the desired stimulus/response pairing (if it is possible for such a pairing to be learned).

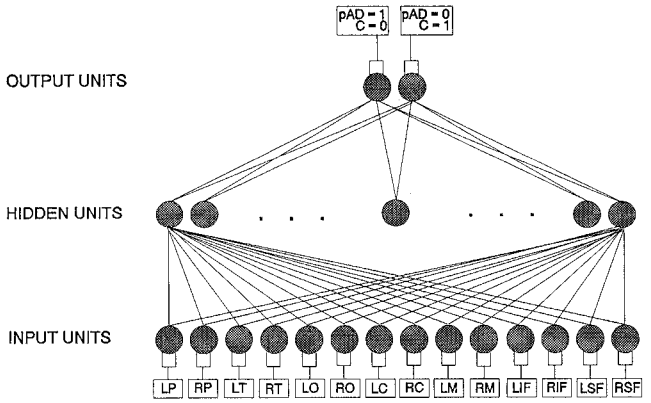

Fig. 1. An illustration of an artificial neural network architecture. The *circles* represent processing units, and the *lines between circles* represent modifiable connections. This particular network was used to classify subjects on the basis of SPET scores (see Materials and methods). As indicated by the labels at the bottom, each input unit is used to encode a subject's measure of perfusion for a particular ROI. The two output units are used to identify subject types. As indicated by the labels, one of these units was trained to generate a response of 1 to a pAD patient and to generate a response of 0 to a comparison subject. The other unit was trained to generate the opposite set of responses. A varying number of hidden units were used in different networks, ranging from zero (there were only direct connections between input and output units) to 20

Regression vs ANNs. In order to see why an ANN may be more capable than standard statistical techniques of discriminating pAD from control SPET images, let us briefly describe multiple regression in the ANN framework. Figure 2a represents multiple linear regression as a simple ANN. The input units are used to represent the values of predictor variables, the weights of the connections are used to represent the regression coefficients associated with these variables, and the line in the output unit indicates that the response of the network is only a linear combination of the (weighted) predictors. Figure 2b represents multiple non-linear regression (e.g. probit analysis) as an ANN. The only difference between it and Fig. 2a is the non-linear transformation of (weighted) predictors that is accomplished by the non-linear function in the output unit. It has been proved [20] that the networks illustrated in Fig. 2a and b can only solve linearly separable pattern recognition problems, and that they cannot solve linearly non-separable problems. This is because the activation function in the output unit for either type of network can only be used to carve a single cut through a pattern space.

Figure 2c represents a network from modern connectionism. In contrast to the networks illustrated in Fig. 2a and b, it has a layer of hidden units that permit a variety

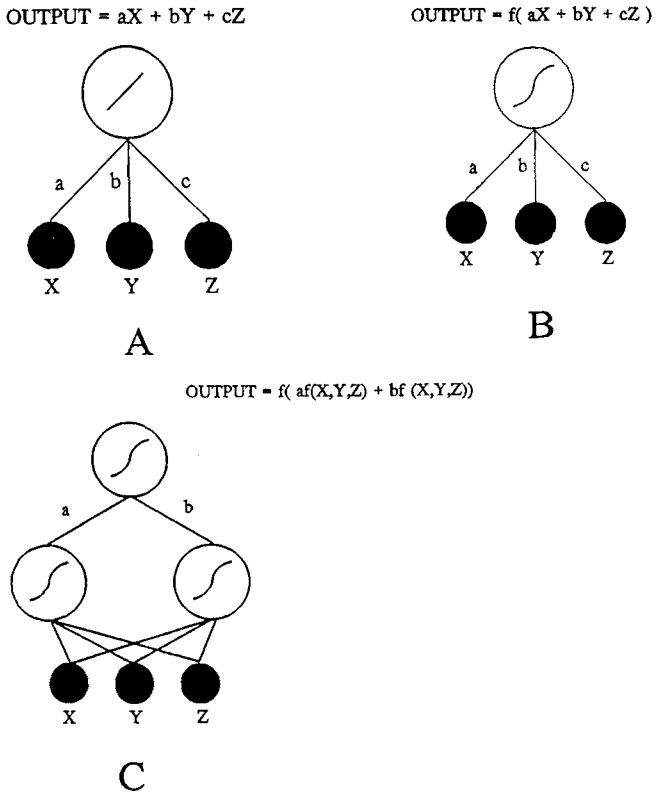

Fig. 2 A–C. A comparison of statistical techniques from a neural network perspective. **A** Multiple linear regression can be viewed as a network with a linear activation function (straight line in output unit) whose connection weights are analogous to coefficients in a regression equation. **B** Multiple non-linear regression is similar, with the exception that some non-linear function *f* defines output unit activity. **C** A modern connectionist network has an additional layer of non-linear feature detectors that result in additional predictive power

of non-linear transformations and combinations of the input variables to be computed prior to their being passed to the output unit. The presence of these hidden units permits such networks to discriminate patterns that could not be discriminated by the other networks because each hidden unit can provide a new partitioning of the pattern space [19, 21].

Purpose of the study

The previous section has argued that, in principle, ANNs are capable of providing better diagnosis of pAD from SPET measurements. Is this true in practice? The research reported below attempted to determine the capability of ANNs to discriminate pAD patients from a healthy comparison group on the basis of SPET. In contrast to many previous studies, a very large and heterogeneous group of pAD patients was examined, to accurately reflect the potential complexity of this pattern recognition task. The diagnostic ability of a number of ANNs, each differing from the others in terms of the number of their hidden units, was assessed. The performance of the ANNs was compared to the ability of multiple linear regression to predict group membership on the basis of the same set of SPET predictors.

Materials and methods

Subjects

Alzheimer's patients. Ninety-seven patients with a diagnosis of pAD consistent with the NINCDS-ADRDA [22] criteria participated in the research. The CAMDEX [23] served as the primary diagnostic tool, with CT scans being obtained for most patients. The CT scans showed no evidence of vascular events or space-occupying lesions.

The patients ranged in age from 63 to 93 years (mean: 78.6 years), with 23 males and 74 females in the group. These patients came from two sources. One set was in-patients recruited from the Geriatric Psychiatry Unit, Edmonton General Hospital. These patients were in the hospital for assessment and rehabilitation, with their average stay being about 4 weeks. The other participants were out-patients who had been referred to the Memory Clinic of the same hospital. There were differences in the male to female ratio (6/42 for the in-patients and 17/32 for the out-patients), but there was little in the way of differences between the two groups in terms of age or dementia severity. The mean age of the in-patient group was 79.2 years and the mean age of the out-patient group was 76.6 years. The number of persons in each of the categories of dementia severity were minimum = 13, mild = 18, moderate = 15 and severe = 2 for the in-patient group, and minimum = 5, mild = 33, moderate = 9 and severe = 2 for the out-patient group.

Comparison subjects. Sixty-four community-dwelling adults served as a comparison group. This group consisted of 25 males and 39 females. The mean age of the group was 53.2 years, and the mean number of years of education for the group was 14.4 years. Comparison subjects had been screened for any neurological impairments (e.g. strokes, head injuries, alcoholic blackouts) that could have led to abnormal SPET images. While there was a significant difference in mean age for these two groups, in normal subjects this alone does not typically lead to differences in SPET measurements [24].

Imaging procedures

All individuals were imaged according to the same protocol; the protocol for each group of subjects was approved by the hospital ethics committee and all individual subjects or their caregivers gave appropriate written informed consent prior to the procedure.

Prior to injection, each subject rested supine in a quiet, darkened room for 5–10 min with a butterfly inserted. The subject was then injected with 550–750 MBq ^{99m}Tc-HMPAO without changing the ambience and rested for a further 5 min. Imaging was commenced 30–60 min after injection. Image acquisition was identical between the two groups. All images were acquired for 64 frames with a total 100 K per frame acquired. Acquisition time per frame is usually 20–25 s.

Projection images were acquired with a GE 400AC or GE 400AT rotating gamma camera equipped with a low-energy all-purpose parallel hole collimator and interfaced to a Picker PCS 512 dedicated computer running a TSX operating system. The reconstruction parameters have been previously described [25]. Im-

ages were displayed in sagittal, axial and coronal sections for qualitative evaluation and prior to the quantitative data fitting.

Determination of SPET measures

Data from each subject's SPET examination were analysed using an iterative three-dimensional ROI model [25]. The model consists of seven continguous but non-overlapping volumes in each side of the brain, resulting in 14 regions in total [see 25, Fig. 1]. These are labelled as the (left or right) superior frontal (LSF, RSF), inferior frontal (LIF, RIF), parietal (LP, RP), occipital (LO, RO), temporal (LT, RT), midbrain (LM, RM) and cerebellar (LC, RC) regions. After the model is fit, data from an individual patient are represented as the total number of counts measured in each of the 14 different volumes. For the purpose of analysis by the networks, each of these measures for each subject was then divided by 10 000 in order that they be within the range of 0 to approximately 12. This did not alter the relationships of any of the measures to one another, but merely represented these measures in a scale that is typically presented to a network.

Results

Standard statistical analyses

Prior to analysing the SPET data with ANNs, we attempted to determine whether a more traditional statistical approach would suffice. First, descriptive statistics were employed to determine whether there were any ROI measures that revealed significant differences between the two groups. Table 1 provides the mean counts for each ROI for the two groups, as well as the results of independent t tests used to compare the statistical difference between these means. This table reveals four important findings. First, the mean number of counts in a ROI were always lower for the Alzheimer's group than for the comparison group. Second, an examination of the means and standard deviations for each ROI indicates a substantial overlap between the two groups in the distributions of their SPET measures. Third, and not surprisingly given the overlapping distributions, differences between groups were statistically significant for only two ROIs: the left and right midbrain. Fourth, while this study indicates that both temporal and parietal counts were lower for the pAD group than for the control, this difference was not statistically significant.

These last two findings are inconsistent with those of previous researchers. On the one hand, the significant difference between the groups for the two midbrain ROIs are likely an artefact of the dataset. Specifically, in Hooper et al.'s [25] three-dimensional ROI model, the two midbrain volumes are substantially smaller than the other 12 ROIs. The result is lower mean cerebral perfusion, and more importantly lower within-group variability, for these two small ROIs (see Table 1). With the large number of subjects in our sample, this decrease in variability was sufficient to make these two mean differences statistically significant. On the other hand, the failure to find significant differences in the temporal or parietal ROIs is due to the fact that in these larger ROIs, particularly with a large and heterogeneous sample of subjects, within-group variability an overlap between the distributions of the two groups are both high enough to prevent significant between-group differences from emerging. Furthermore, the majority of our patient sample was in the minimum to mild range (71%) and Frlich et al. [14] have shown that dysfunction in temporal and parietal areas is correlated with dementia severity.

Our second statistical approach to this dataset was to use multiple regression to determine whether a linear combination of the 14 SPET measures could be used to reliably discriminate the two groups. In this analysis, the predicted variable was group membership, which was coded as −1 for the pAD patients and as 1 for the comparison group. The regression equation that included all 14 SPECT measures was statistically significant (R = 0.579, F = 5.323, df = 14, 148; $P < 0.001$). Even so, this equation still accounted for only 33.5% of the variance in the data.

Table 1. Mean scaled counts for the two groups as measured in each region of interest; standard deviations of each mean are presented in parentheses. The t tests used to compare means used pooled variances, and each had 159 degrees of freedom

Region of interest	Perfusion in ROI		t	Probability
	p^{AD}	Comparison		
RSF	5.001 (1.269)	5.241 (1.264)	−1.188	0.237
LSF	4.960 (1.246)	5.138 (1.267)	−0.892	0.373
RIF	5.247 (1.314)	5.422 (1.330)	−0.831	0.407
LIF	5.198 (1.318)	5.331 (1.297)	−0.638	0.525
RP	11.761 (2.967)	12.279 (2.925)	−1.099	0.273
LP	11.802 (2.968)	12.187 (2.935)	−0.816	0.415
RO	11.038 (2.815)	11.623 (2.747)	−1.316	0.190
LO	11.107 (2.864)	11.645 (2.692)	−1.205	0.230
RT	11.268 (2.832)	11.776 (2.727)	−1.142	0.255
LT	11.245 (2.807)	11.690 (2.754)	−1.000	0.319
RM	3.214 (0.785)	3.550 (0.834)	−2.621	0.010**
LM	3.199 (0.778)	3.551 (0.842)	−2.746	0.007**
RC	7.195 (1.750)	7.408 (1.548)	−0.798	0.426
LC	7.375 (1.771)	7.545 (1.612)	−0.622	0.535

In summary, while traditional statistical analyses did reveal a few statistically significant effects, they did not provide sufficient predictive power to serve as a diagnostic instrument.

Results from ANNs

A series of ANNs were trained to discriminate the two groups of subjects using the SPET data as input. Each ANN had 14 input units, each of which was used to encode counts for a particular ROI (see Fig. 1). Each ANN also had two output units. One of these output units was trained to generate a response of 1 to a pAD patient and a response of 0 to a comparison subject. The other output unit was trained to generate a response of 1 to a comparison subject, and a response of 0 to a pAD patient. We used two output units instead of one to allow the network to encode anomalies; for example, if both output units generated 0 activation, then this would indicate that the subject in question was not characteristic of the Alzheimer's group, but was also not characteristic of the comparison group. Networks differed from one another only in terms of the number of hidden units that were employed. We examined five different types of networks: 0 hidden units (i.e. there were direct connections from input to output), and 5, 10, 15 and 20 hidden units. This approach was taken because the predictive power of a network (i.e. its ability to learn the desired discrimination) should increase with the number of hidden units employed.

Dawson and Schopflocher's [10] value unit architecture was used to define the properties of each processing unit in the networks. In this architecture, every output unit and hidden unit uses a Gaussian activation function to perform a non-linear transformation of its net input. As a result, each hidden unit makes two cuts in a pattern space, instead of the single cuts provided by the units illustrated in Fig. 2. This architecture was adopted because previous studies have shown that it can outperform standard architectures [e.g. 19] at difficult pattern discrimination tasks. Furthermore, there is reason to believe that value unit networks are better at generalizing what they have learned to new instances [13], and that it is easier to interpret how a particular network is performing a particular pattern recognition task [26].

Each network was trained using Dawson and Schopflocher's [10] elaboration of Rumelhart et al.'s [19] generalized delta rule, with a learning rate of 0.03 and a momentum of 0. Connection weights and unit biases (i.e. the mean of the Gaussian) were randomly selected from the range −0.3 to +0.3. Each network was trained for 20 000 epochs (i.e. 20 000 presentations of each subject's data); the order of pattern presentation was randomized during learning. At the end of training, the response of each output unit in the network was determined for each subject's SPET data, and the proportion of variance accounted for was used as a measure of the network's diagnostic ability. (Note that each output unit is being trained to discriminate all 161 subjects, and as a result proportion of variance accounted for can be computed independently for output unit 0 and output unit 1.) In order to determine the average ability of the networks to learn this pattern discrimination task, ten separate training sessions were conducted for each size of network.

The mean proportion of variance accounted for by the networks is illustrated in Fig. 3. Two conclusions can be drawn from this figure. First, on average the ANN accounted for a great deal more of the variance in the data than is accounted for by the multiple regression. Second, as the number of hidden units is increased, the predictive power of the networks also increases. Indeed, if one uses 15 or 20 hidden units, then one is accounting for between 80% and 90% of the variance in the data.

While none of the network that were trained accounted for 100% of the variance of the data, this does not necessarily mean that the networks were unable to perfectly discriminate the two groups of subjects. Figure 4 presents a scatterplot of the output activities produced by each subject's SPET data in a 15 hidden unit network. It can be seen that in this scatterplot there is absolutely no overlap between members of the two groups of subjects. Clearly, even with this sample of relatively mild pAD patients, the ANN approach was able to differentiate the

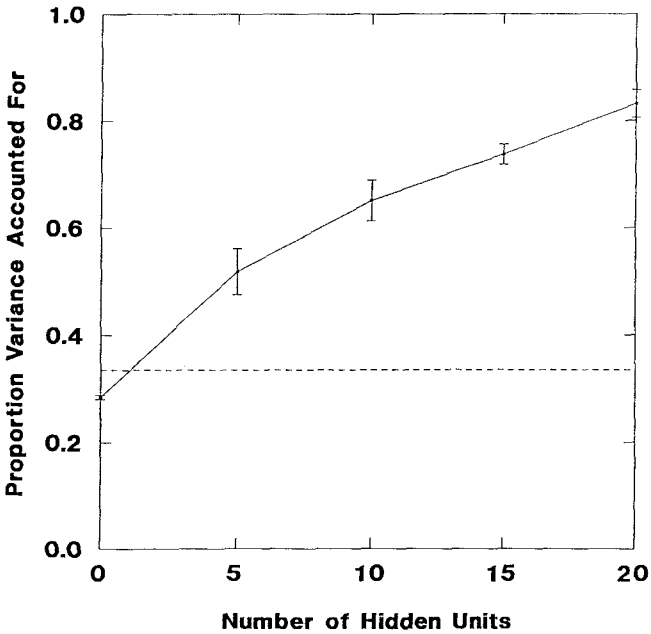

Fig. 3. Diagnostic power of the network using proportion of variance accounted for as a measure of the fit between the actual output of an output unit and the desired output of the unit. The *dashed line* represents the proportion of variance accounted for by the multiple regression equation. Note that the *solid line* plots the mean proportion of variance accounted for (with standard error) by unit 0, which was being trained to generate a 1 for a pAD subject and a 0 for a comparison subject. Nearly identical results are obtained for unit 1, and for average overall network performance (not plotted)

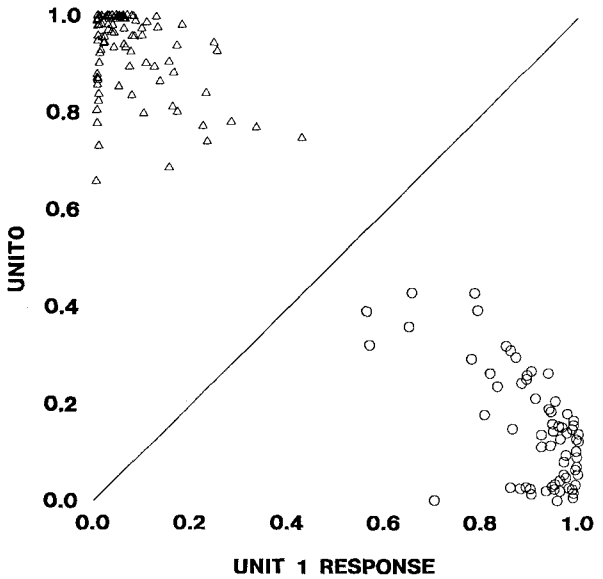

Fig. 4. A two-dimensional scatterplot showing the network's discrimination of 161 subjects by combining unit 0 and unit 1 outputs. In this plot, *triangles* represent pAD subjects and *circles* represent comparison subjects. A *diagonal line* has been drawn through this scatterplot to show that the network has discriminated these two groups of subjects perfectly

patient and normal samples when the traditional statistical approach was unsatisfactory.

Transfer of learning to new cases

The goal of the simulations described above was to determine the ability of ANNs to discover a mapping between SPET measures and patient categories. In light of this goal, the networks were trained on a single data set, and the dependent measure reflected the amount of classification error that remained at the end of training. The results of these simulations demonstrated that ANNs are very capable of learning a mapping that allows accurate patient classification on the basis of the input data.

Some researchers, with their first glance at Figs. 3 and 4, might argue that these results are not interesting. After all, during training the network is "told" which patient category each set of SPET scores belongs to. As a result, these researchers tend not to be surprised that the network can learn to categorize patients in the training set. In their mind, it would be much more interesting to see how well a network transfers what is has learned about one set of patients to new patients who have not previously been used during training.

This view of the data fails to recognize the fact that discovering an accurate mapping from SPET measures to patient categories is a non-trivial achievement. For example, the multiple regression technique to which the ANNS were compared also "knew" which patient category each set of SPET scores was associated with, but was far less successful in using this information to determine an accurate metric for subject classification. In our view, one should determine that a system has the potential to make accurate diagnoses *prior* to testing its ability to generalize. Testing the ability of a system to generalize when it was, say, only 25% accurate on a training set would be fruitless. This perspective is particularly germane when standard statistical techniques are relatively poor in using the SPET measures for diagnosis.

The results reported above clearly indicate that ANNs have the potential to be accurate tools for the diagnosis of pAD on the basis of SPET measures. As a result, the next logical step in evaluating their potential for application in a clinical setting is to investigate how well performance generalizes to new patient samples.

ANNs are attractive because they have, in principle, very high potential to generalize what they have learned to new instances [e.g. 27]. In practice, however, the ability of an ANN to transfer its knowledge is strongly affected by many factors, including the number of hidden units in the network, the number of connections in the network, the amount of training sweeps that have been used, the number of patterns in the training set, the types of processing units that are used in the network and the complexity of the mapping that the network is learning [e.g. 13, 28–30]. As a result, a complete study of the factors that produce good (or bad) generalization in the networks described above is beyond the scope of the current manuscript. However, pilot results on network generalization have been very encouraging.

For example, in one study we randomly deleted 40 patients from the training set, and then trained a 15 hidden unit network on the remaining patients using the same methods described above. At the end of training, the network's ability to categorize the 40 patients that it had not seen was tested. This was done by recording the network's response to each patient's set of SPET measures, converting the network's response into a forced choice diagnosis. This conversion was accomplished by assigning the subject to the patient category to which the network's response was more similar. (In graphical terms, if the network's response was plotted as a point in Fig. 4, then the patient was assigned to the patient category to which this point was closest to. In Fig. 4 the Alzheimer category is the upper left hand corner, and the comparison category is the lower right hand corner.) The results of this simulation indicated that the network achieved an overall accuracy rate of 72.5% in classifying the 40 new patients (18 correct classifications of 24 pAD patients and 11 correct classifications of 16 comparison patients). We are currently studying generalization more rigorously in order to determine how changes in network architecture (e.g. reductions in number of hidden units, reduction of the amount of training of the network) can produce even higher transfer rates.

Discussion

While a few researchers have attempted to use ANNs to model to memory deficits observed in Alzheimer's disease [e.g. 31], to our knowledge no one has attempted to

use ANNs as an aid to its diagnosis. The results presented above clearly indicate that ANNs have a significant amount of clinical potential in this field. As well, these results confirm those of previous researchers, and indicate that SPET also offers an extremely valuable tool for the early diagnosis of this disorder.

However, the results also indicate three related caveats to the clinical use of SPET for this disease. First, our "traditional" statistical analyses indicated that even though one could find statistically significant differences between group means, these differences could not be translated into a powerful quantitative measure capable of discriminating the two groups of subjects. Second, a single positive feature for the diagnosis of Alzheimer's disease did not reveal itself. Third, as subject samples become larger and more heterogeneous, typical indices used to discriminate pAD patients from other groups – such as reduced parietal and temporal lobe perfusion – clearly do not work as well. In short, our data suggest that pAD diagnosis using SPET is a linearly non-separable pattern recognition problem. As a result, ANNs, or some other sophisticated statistical technique capable of detecting complex non-linear relationships, may be necessary to tap the diagnostic potential of SPET.

While these results are extremely encouraging, many additional questions must still be addressed. Our current research is focusing on three different issues: the extent to which the predictive power of the current networks will generalize to new samples, the extent to which a network can discriminate more than two different groups of subjects on the basis of SPET, and the extent to which one can interpret the structure of a trained network to determine the kinds of non-linear relationships being used to determine group membership.

Acknowledgements. This research was supported by NSERC Research Grant A2038, NSERC Equipment Grant 138704 and a University of Alberta Central Research Fund grant, all awarded to M.R.W.D., by a National Health Research and Development Program grant awarded to A. Dobbs and B. Rule, and by a Medical Services Incorporated grant awarded to A. McEwan, Z. Catz, A. Dobbs, B. Rule, K. Rockwood and J. Triscott. We would also like to thank Barbara Carstensen, Collene Lust, Diedre Gailey, Charlene Pisio, Bob Heller, Liz Panasiuk and Matthew Duncan for their assistance with this project.

References

1. Battistin L, Pizzolato G, Dam M, Ponza I, Borsato N, Zanco PL, Ferlin G. Regional cerebral blood flow study with 99mTc-hexamethyl-propylenamine oxime single photon emission computed tomography in Alzheimer's and multi-infarct dementia. *Eur Neurol* 1990; 30: 296–301.
2. Burns A, Philpot MP, Costa DC, Ell PJ, Levy R. The investigation of Alzheimer's disease with single photon emission tomography. *J Neurol Neurosurg Psychiatry* 1989; 52: 248–253.
3. Gemmell HG, Sharp PF, Besson JAO, Ebmeier K, Smith FW. A comparison of Tc-99m-HM-PAO and I-123 IMP cerebral SPET images in Alzheimer's disease and multi-infarct dementia. *Eur J Nucl Med* 1988; 14: 463–466.
4. Hunter R, McLuskie R, Wyper D, Patterson J, Christie JE, Brooks DN, McCulloch J, Fink G, Goodwin GM. The pattern of function-related regional cerebral blood flow investigated by single photon emission tomography with 99mTc-HMPAO in patients with presenile Alzheimer's disease and Korsakoff's psychosis. *Psychol Med* 1989; 19: 847–855.
5. Jagust WJ, Budinger TF, Reed BR. The diagnosis of dementia with single photon emission computed tomography. *Arch Neurol* 1987; 44: 258–262.
6. Neary D, Snowden JS, Shields RA, Burjan AWI, Northen B, Macdermott N, Prescott MC, Testa HJ. Single photon emission tomography using 99mTc-HM-PAO in the investigation of dementia. *J Neurol Neurosci Psychiatry* 1987; 50: 1101–1109.
7. Schmitt FA, Shih W, DeKosky SJ. Neuropsychological correlates of single photon emission computed tomography (SPECT) in Alzheimer's disease. *Neuropsychology* 1992; 6: 159–171.
8. Smith FW, Besson J, Gemmell H, Sharp P. The use of technetium-99m-HM-PAO in the assessment of patients with dementia and other neuropsychiatric conditions. *J Cereb Blood Flow Metab* 1988; 8: 5116–5122.
9. Albert MS, Lafleche G. Neuroimaging in Alzheimer's disease. *Psychiatr Clin North Am* 1991; 14: 443–459.
10. Dawson MRW, Schopflocher DP. Modifying the generalized delta rule to train networks of nonmonotonic processors for pattern classification. *Connection Sci* 1992; 4: 19–31.
11. Pao Y. *Adaptive pattern recognition and neural networks.* Reading, Mass. Addison-Wesley.
12. Ahmad S, Tesauro G. Scaling and generalization in neural networks: a case study. In: Hinton GE, Sejnowski TJ, Touretzky DS (eds) *Proceedings 1988 connectionist models summer school.* San Mateo Calif. Morgan Kaufmann; 1988.
13. Shamanski KS, Dawson MRW, Berkeley ISN. The effect of linear separability on learning speed and generalization in monotonic and nonmonotonic PDP networks. *Connection Sci* 1994; in press.
14. Frlich L, Eilles C, Ihl R, Maurer K, Lancik M. Stage-dependent reductions of regional cerebral blood flow measured by HMPAO-SPECT in dementia of Alzheimer type. *Psychiatry Res* 1989; 29: 347–350.
15. Dobbs A, Heller R, McEwan A. *What is the future of staging in dementia? Patient variability.* Symposium on Dementia, Canadian Association on Gerontology, Toronto.
16. Martin A, Brouwers P, Lalonde F, Cox C, Teleska P, Fedio P. Towards a behavioral typology of Alzheimer's patients . *J Clin Exp Neuropsychol* 1986; 8: 594–610.
17. McClelland JL, Rumelhart DE, Hinton GE. The appeal of parallel distributed processing. In: Rumelhart DE, McClelland JL and the PDP Research Group (eds) *Parallel distributed processing, V.1.* Cambridge, Mass.: MIT Press; 1986.
18. Smolensky P. On the proper treatment of connectionism. *Behav Brain Sci* 1988; 11: 1–74.
19. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature* 1986; 323: 533–536.
20. Minsky M, Papert S. *Perceptrons, 3rd edn.* Cambridge, Mass.: MIT Press.
21. Lippmann RP. An introduction to computing with neural nets. *IEEE ASSP Magazine* 1987; April: 4–22.
22. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Statian EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of the Dept. of Health and Human Sources Task Force on Alzheimer's Disease. *Neurology* 1984; 34: 939–944.

23. Roth M, Huppert TA, Tym E, Mountjoy CQ. *The Cambridge examination for mental disorders of the elderly.* Cambridge: Cambridge University Press, 1988.
24. McEwan AJB, Hooper HR, Kotchon T, McTeer P, Hooper PM. A three dimensional programme to quantitatively analyze cerebral perfusion from HMPAO images: normal and pathophysiological data. *Nucl Med Commun* 1990; 11: 221.
25. Hooper HR, McEwan AJ, Lentle BC, Kotchon TL, Hooper PM. Interactive three-dimensional region of interest analysis of HMPAO SPECT brain studies. *J Nucl Med* 1990; 31: 2046–2051.
26. Berkeley ISN, Dawson MRW, Medler DA, Schopflocher DP. Density plots of hidden value unit activations reveal interpretable bands and microband. Neural Computation 1994; in press.
27. Hinton GE, McClelland JL, Rumelhart DE. Distributed representations. In: Rumelhart DE, McClelland JL (eds) *Parallel distributed processing, vol. 1.* Cambridge, Mass: MIT Press; 1986: 77–109.
28. Baldi P, Chauvin Y. Temporal evolution of generalization during learning in linear networks. *Neural Computation* 1991; 3: 589–603.
29. Hertz J, Krogh A, Palmer RG. *Introduction to the theory of neural computation.* Redwood City, Calif. Addison-Wesley, 1991.
30. Baum EB, Haussler D. What size net gives valid generalization? *Neural Computation* 1989; 1: 151–160.
31. Horn D, Ruppin E, Usher M, Herrmann M. Neural network modeling of memory deterioration in Alzheimer's disease. *Neural Computation* 1993; 5: 736–749.