What can artificial neural networks teach us about neurodegenerative disorders with extrapyramidal features?


Summary
Artificial neural networks (ANNs), computer paradigms that can learn, excel in pattern recognition tasks such as disease diagnosis. Artificial neural networks operate in two different learning modes: supervised, in which a known diagnostic outcome is presented to the ANN, and unsupervised, in which the diagnostic outcome is not presented. A supervised learning ANN could emulate human expert diagnostic performance and identify relevant predictive markers in the diagnostic task, while an unsupervised learning ANN could suggest reasonable alternative diagnostic classification criteria. In the present study, we used ANN methodology to try to overcome the neuropathological difficulties in differentiating the subtypes of progressive supranuclear palsy (PSP), and in differentiating PSP from postencephalitic parkinsonism (PEP) and corticobasal degeneration, or Pick's disease from corticobasal degeneration. First, we applied supervised learning ANN to classify 62 cases of these disorders and to identify diagnostic markers that distinguish them. In a second experiment, we used unsupervised learning ANN to investigate possible alternative nosological classifications. Artificial neural networks input data for each case consisted of values representing histological features, including neurofibrillary tangles, neuronal loss and gliosis found in multiple brain sampling areas. The supervised learning ANN achieved excellent accuracy in classifying PSP but had difficulty classifying the other disorders. This method identified a few features that might help to differentiate PEP, supported currently proposed criteria for Pick's disease, corticobasal degeneration and typical PSP, but detected no features to characterize the atypical subtype of PSP. In general, unsupervised learning ANN supported the present nosological classification for PSP, PEP, Pick's disease and corticobasal degeneration, although it overlapped some groups. Artificial neural networks methodology appears promising for studying neurodegenerative disorders.

Keywords: artificial neural networks; dystal; neurodegenerative disorders; neuropathology; progressive supranuclear palsy

Abbreviations: ANN = artificial neural network; PEP = postencephalitic parkinsonism; PSP = progressive supranuclear palsy

© Oxford University Press 1996
Introduction

Although neuropathological examination provides essential information for diagnosing neurodegenerative disorders with extrapyramidal features, neuropathologists may have difficulty distinguishing subtypes of PSP, PEP, corticobasal degeneration and Pick's disease (Steele et al., 1964; Geddes et al., 1993; Hauw et al., 1994). These disorders have not only common clinical symptom profiles, but also common histopathological features such as neurofibrillary tangles, gliosis and neuronal loss. We recently confirmed that it is nearly impossible to distinguish PEP from PSP using only neuropathological information (Steele et al., 1964; Geddes et al., 1993; Litvan et al., 1996) and found severe difficulties in differentiating the proposed subtypes of PSP neuropathologically (Hauw et al., 1994; Lantos, 1994). Similarly, some cases of corticobasal degeneration are difficult to distinguish from PSP because basophilic inclusions in corticobasal degeneration may appear identical to neurofibrillary tangles in PSP (Litvan et al., 1996). Additional confusion arises in the neuropathological differentiation of corticobasal degeneration, Pick's disease and PSP. Achromatic neurons in corticobasal degeneration share many morphological and immunocytochemical characteristics with swollen cells in Pick's disease, and swollen cells may be present in PSP. Moreover, cases of corticobasal degeneration, Pick's disease and PSP may overlap neuropathologically (Seitelberger et al., 1983; Arima et al., 1992; Daniel et al., 1994).

Such complexities prompted us to explore the use of ANNs, which reportedly excel in pattern recognition and classification tasks such as disease diagnosis (Dayhoff, 1990; Hecht-Nielsen, 1990). Artificial neural networks are computing paradigms inspired by the neurosciences (Hebb, 1949; Levine, 1990). These paradigms are implemented in both software and hardware; however, most applications to date have employed software (i.e. computer programs). These programs are not algorithmic in the sense that they would classify cases according to some pre-programmed rules or criteria, but rather they are 'machine-learning programs' that learn diagnostic classifications through a training process that consists of repeated exposure to data representing exemplar cases.

Artificial neural networks can be designed to operate in two different learning modes: supervised, in which correct diagnostic outcomes are presented to the ANN during training, and unsupervised, in which diagnostic outcomes are not presented. A supervised learning ANN could learn to emulate human expert diagnostic performance and may identify the important predictive markers that human experts use in performing diagnostic tasks. Supervised learning ANNs operate in two phases: a learning or training phase, in which cases are presented to the ANN with correct diagnostic information and a validation phase, in which new cases are presented to the ANN with known diagnostic information withheld. Performance is evaluated in the validation phase by comparing the ANN predicted diagnoses with the known diagnoses of the new cases. In unsupervised learning, the ANN is not provided with diagnostic outcomes, and it clusters cases according to some preselected rules of similarity, thus possibly suggesting alternative diagnostic classification criteria.

Artificial neural networks have been used successfully in the past decade for diagnostic classification in a wide range of medical problems (Baxt, 1991; Nafe and Choritz, 1992; Reggia, 1992; Litvan et al., 1993). For example, in one study (Baxt, 1991), an ANN successfully identified 97% of patients with acute myocardial infarction (sensitivity) and 96% without infarction (specificity), which was comparable to a sensitivity of 78% and a specificity of 85% achieved by human experts. When used with pathological data in the diagnosis of thyroid tumours (Nafe and Choritz, 1992), ANN methods were accurate and avoided separate classifications for tumours not considered significantly different.

In neurology, ANNs have been used for the diagnosis of dementia and epilepsy, and for neurological localization (Mulsant and Servan-Schreiber, 1988; Apolloni et al., 1990; Cho and Reggia, 1993; Anderer et al., 1994; Kippenhan et al., 1994; Tuhrim et al., 1994; DeFigueiredo et al., 1995). In dementia, ANN methodology achieved 90% correct classification of demented and control subjects when applied to fluorodeoxyglucose PET images (Kippenhan et al., 1994) or topographic EEG slow-wave activity (Anderer et al., 1994). Artificial neural networks also discriminated between clinically diagnosed groups of elderly normal subjects and patients with Alzheimer's disease or vascular dementia in a single photon emission CT study with technetium-99 images (DeFigueiredo et al., 1995). In epilepsy, ANN methods applied to the classification of types of epilepsy, with signs and symptoms as input data, were 80% correct for single diagnoses and 95% correct for clusters of diagnoses (Apolloni et al., 1990). Regarding neurological localization, an ANN precisely determined the site of brain damage in stroke patients (Tuhrim et al., 1994).

There have been no reports on the application of ANN methods to study neuropathological diagnostic classifications of neurodegenerative disorders with extrapyramidal features. Our aims were to investigate the accuracy with which an ANN could classify these disorders, its ability to identify features that could distinguish them, and to explore possible ANN-suggested alternative nosological classifications. For these purposes we designed two experiments. In the first experiment, we applied supervised learning ANN methodology to classify the neurodegenerative disorders with extrapyramidal features and to identify the relevant neuropathological diagnostic markers that might distinguish them. In the second experiment, we used unsupervised learning ANN methodology to investigate ANN-suggested groupings of these disorders as possible alternatives to those proposed by human experts. Data consisted of values representing
neuropathological features from cases of neurodegenerative disorders with extrapyramidal features provided by experienced neuropathologists from their own research and clinical files.

We chose to work with the dynamically stable associative learning (Dystal) ANN paradigm developed by Alkon and colleagues (Alkon, 1989; Alkon et al., 1990, 1994) because it is suitable for conducting both supervised and unsupervised learning studies, and because we had experience with using it. Computer models and software for conducting the supervised and unsupervised Dystal ANN experiments for this study were designed, developed, and executed by one of the authors (J.M.D.).

Methods
The data were provided by eight neuropathologists from centres specializing in the study of neurodegenerative disorders, who selected a total of 62 cases from their own research and clinical files. The sample included 18 typical PSP cases, nine atypical PSP cases, four combined PSP cases, 12 corticobasal degeneration cases, eight PEP cases and 11 Pick’s disease cases, all of which fulfilled the recently proposed preliminary National Institutes of Neurological Disorders and Stroke (NINDS) criteria (Hauw et al., 1994). Cases had to include all but one of the following sampling areas: globus pallidus, putamen, caudate nucleus, subthalamic nucleus, midbrain, pons, medulla, dentate nucleus of the cerebellum, hippocampus, parahippocampal gyrus, and motor, frontal or parietal cortices; and at least two of the following staining: haematoxylin-eosin and either a silver impregnation (modified Bielschowsky, Bodian, Gallyas and its modifications) or tau and ubiquitin immunohistochemistry (Hauw et al., 1994). However, values corresponding to parietal, temporal and motor cortices that had >20% of the data missing were not included in the analysis.

The same eight neuropathologists independently evaluated their own cases, and immediately thereafter filled out a standardized form that listed the pathological features potentially present in any of the selected disorders. The neuropathologists had agreed on the definition of the pathological features and how to score them. The morphological features chosen for assessment included neurofibrillary tangles and neuronal loss in the previously described brain cortical and subcortical sampling areas: presence of swollen cells (achromatic cell), Pick bodies, argyrophilic tau process, Lewy bodies, corticobasal inclusions and neuritic plaques. Scoring of lesions was semiquantitative (scale of 0–2, with 0 being none, 1 mild and 2 severe). The neuropathologist’s judgment on the importance of macroscopy and histology in the assessment of each case was also recorded (scale of 1–5, with 1 being the most important feature and 5 the least important feature). The definitive diagnosis for each case was taken to be the one provided by the neuropathologist, who examined his/her

The same eight neuropathologists independently evaluated their own cases, as they used all the available clinical and pathological records.

Experiment 1: supervised learning ANN
The training data for input to the ANN consisted of values for 76 putative predictive variables. These values represented the morphological features provided by the neuropathologists and the diagnostic outcome for each case (definite diagnosis). The training data set was presented to the input layer of the neural network, one case at a time (Fig. 1). Data representing each case were then transferred to the hidden layer where it was compared for ‘similarity’ to ‘patches’ constructed with data from previously processed cases with the same diagnostic outcome. The patch construction process employed a similarity measure to cluster cases in diagnostic classes and to associate each cluster with one diagnosis node in the output layer. When the first training case was presented, there were no patches in the hidden layer; thus, the first case was used to construct the first patch and its corresponding output diagnosis node. When each subsequent training case was presented, it was compared for similarity to each existing patch having the same diagnostic outcome. If none were found, the new case was used to construct a new patch and associated output node. When a new training case was similar to an existing patch, the new case was mathematically ‘merged’ with that patch. This merging produced a new patch with 76 individual data elements containing the mean data element values of all the cases represented by that patch. Dystal can detect multiple distinct clusters corresponding to the same disease. Thus, it is possible to have more than one output node corresponding to an established diagnostic class. This feature of Dystal allows cases with very different feature profiles to share the same diagnostic classification.

In the validation phase, new cases were presented without diagnostic outcome. Patches established during training were used for validation. The input layer sequentially presented
Table 1: Accuracy of the diagnosis of neurodegenerative disorders with extrapyramidal features

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Positive predictive value</th>
<th>Relative frequency†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dystal supervised ANN*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSP</td>
<td>0.87</td>
<td>0.94</td>
<td>0.93</td>
<td>0.50</td>
</tr>
<tr>
<td>Atypical PSP</td>
<td>0.67</td>
<td>0.68</td>
<td>0.26</td>
<td>0.15</td>
</tr>
<tr>
<td>Corticobasal degeneration</td>
<td>0.92</td>
<td>0.78</td>
<td>0.50</td>
<td>0.19</td>
</tr>
<tr>
<td>Pick's disease</td>
<td>0.91</td>
<td>0.86</td>
<td>0.59</td>
<td>0.18</td>
</tr>
<tr>
<td>Atypical PSP versus typical and atypical PSP</td>
<td>1</td>
<td>0.85</td>
<td>0.67</td>
<td>0.23</td>
</tr>
<tr>
<td>Neuropathologists†</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSP</td>
<td>0.76</td>
<td>0.85</td>
<td>0.86</td>
<td>0.55</td>
</tr>
<tr>
<td>Atypical PSP</td>
<td>0.14</td>
<td>0.85</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>Corticobasal degeneration</td>
<td>0.75</td>
<td>0.89</td>
<td>0.43</td>
<td>0.09</td>
</tr>
<tr>
<td>Pick's disease</td>
<td>0.91</td>
<td>0.94</td>
<td>0.72</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>0.85</td>
<td>0.97</td>
<td>0.85</td>
<td>0.16</td>
</tr>
</tbody>
</table>

*Total number of cases (n = 62); †total number of observations (n = 163) made by eight neurologists on 62 cases; ‡relative frequency is the proportion of cases of a given disorder among all assigned cases.

the data representing each new case to the hidden layer where it was compared for similarity to all existing patches. The case was assigned to the diagnostic class corresponding to the output node associated with the patch determined to be most similar to the case. Comparing the neural network diagnostic assignments with the known diagnostic outcome produced sensitivity, specificity, and positive predictive values as measures of accuracy (Galen and Gambino, 1975). For a given disorder, sensitivity refers to the proportion of genuine disorder cases correctly classified by the ANN, among all of the genuine disorder cases. Specificity refers to the proportion of genuine cases correctly classified by the ANN as not having the disorder, among all genuine cases not having the disorder. Positive predictive value refers to the proportion of genuine cases of the disorder among all cases classified as having the disorder by the ANN.

Since our data set was small, the leave-one-out, or jackknife method of training and validation was applied (Efron, 1982; Efron and Tibshirani, 1993). With the jack-knife, each case was individually classified into a diagnostic category (i.e. validated) by a unique ANN trained with all of the other 61 remaining cases.

Seven studies were conducted using the Dystal supervised learning ANN with the 76 putative covariate neuro-pathological data values representing the 62 cases and the five diagnostic categories. In each of the first five studies, data sets were constructed for each diagnostic category in which each case was either a member or a nonmember of that category (e.g. PSP versus non-PSP). To further investigate the diagnostic difficulties between PEP and PSP, the sixth study was conducted with a data set that contained only cases with these diagnoses (PEP and typical and atypical PSP). In the seventh study, the ANN was applied to the full data set in order to identify morphological features that may distinguish all the evaluated disorders.

Experiment 2: unsupervised learning ANN
For Experiment 2, we studied 58 of the 62 cases described in the first experiment, excluding the four cases with combined disorders. The entire 58-case data set was used for training, and there was no validation phase. The training data set had the same format as described earlier for the first experiment, but the diagnostic outcome was excluded. Each case presented to the ANN contained values for the 76 data variables. Cases were sequentially presented to the input layer of the neural network (Fig. 1). Data representing each case were transferred to the hidden layer where it was compared with existing patches for similarity. Data for the first case formed the first patch, since initially there were no patches in the hidden layer. Each subsequent case was evaluated for similarity to existing patches. If a new case was not similar to any existing patches, the case was used to create a new patch and a new diagnostic category. If a new case was similar to one or more patches, the case clustered with the most similar patch. When a case was added to a pre-existing patch, mean values of all 76 covariates were updated in that particular patch. The ANN classifies cases according to built-in similarity rules that employ a variable ‘similarity threshold’ that ranges from 0 to 1. Changing the similarity threshold alters the manner in which cases are clustered. If the similarity threshold is high, much attention is paid to detail, and the ANN could produce a new patch for each case presented (e.g. 58 in this case). If the similarity threshold is low, the ANN could cluster all cases into a single patch.

Results

Experiment 1
The Dystal supervised learning ANN produced the sensitivity, specificity and positive predictive value measurements shown...
Table 2  Features identified by the ANN for differentiating disorders

<table>
<thead>
<tr>
<th>Disorder differentiation</th>
<th>Inclusionary histological features</th>
<th>Exclusionary histological features</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSP from non-PSP</td>
<td>Neurofibrillary tangles in pallidum, putamen, subthalamic nucleus, substantia nigra, periaqueductal grey, tegmentum, basis pontis, inferior olives and dentate; neuronal loss in cranial nerve X</td>
<td>Pick bodies</td>
</tr>
<tr>
<td>Atypical PSP from all other disorders</td>
<td>No features identified</td>
<td>No features identified</td>
</tr>
<tr>
<td>Postencephalitic parkinsonism from all other disorders</td>
<td>No features identified</td>
<td>Swollen cells; argentophilic tau processes in the cortex; neurofibrillary tangles in inferior olives; neuronal loss or gliosis in the putamen, caudate, red nucleus, basis pontis, inferior olives, dentate and nucleus basalis of Meynert</td>
</tr>
<tr>
<td>Postencephalitic parkinsonism from typical and atypical PSP</td>
<td>No features identified</td>
<td>Neurofibrillary tangles in the globus pallidus, locus ceruleus, basis pontis, inferior olives and cranial nerve XII; neuronal loss or gliosis in the globus pallidus, putamen, red nucleus and basis pontis</td>
</tr>
<tr>
<td>Corticobasal degeneration from other disorders</td>
<td>Corticobasal inclusions; swollen cells</td>
<td>Neurofibrillary tangles in the dentate nuclei</td>
</tr>
<tr>
<td>Pick's disease from other disorders</td>
<td>Pick bodies in neocortex and elsewhere; neuronal loss and gliosis in the temporal, frontal lobe and hippocampus; anterior temporal atrophy</td>
<td>Neurofibrillary tangles in the basal ganglia, brainstem and cortex; neuronal loss and gliosis in the subthalamic nucleus, red nucleus, substantia nigra</td>
</tr>
</tbody>
</table>

in Table 1. When trained with the full data set, the ANN identified features that could distinguish PSP from non-PSP cases, PEP from non-PEP cases, PEP from PSP cases, corticobasal degeneration from noncorticobasal degeneration cases, and Pick's disease from other disorders (Table 2). However, it could not identify features to separate atypical PSP from any of the other disorders (i.e. from typical PSP, corticobasal degeneration, PEP and Pick's disease).

Experiment 2

With the similarity threshold set to the mid-range value 0.5, the unsupervised ANN clustered cases into two groups (Table 3). Group 1 contained all the typical PSP cases, 56% of the atypical PSP cases and 25% of the corticobasal degeneration cases. Group 2 contained all cases of PEP and Pick's disease, 75% of corticobasal degeneration cases, and the remaining atypical PSP cases. When the similarity threshold was increased to 0.6, the ANN clustered the cases into six groups (Table 3). One group had 94% of the typical PSP cases, the second group had 100% of the PEP cases and a variable percentage of cases with the other conditions, the third group had 82% of the Pick's disease cases, the fourth group had 67% of the corticobasal degeneration cases, the fifth and sixth groups had the remaining corticobasal degeneration cases. Atypical PSP cases overlapped between groups 1 and 2. Therefore, the ANN clustered cases into four major groups of disorders with several overlapping borderline cases.

Discussion

Experiment 1

The Dystal supervised ANN achieved excellent accuracy measures when classifying PSP but had difficulty classifying the other disorders. The ANN had clear problems predicting the diagnosis of atypical PSP cases. In fact, it had more false positive than true positive diagnoses. Although the ANN achieved high sensitivity for the diagnosis of cases with PEP, corticobasal degeneration and Pick's disease, it obtained both low specificity and positive predictive values for their diagnosis.

In a previous study (Litvan et al., 1996), in which most of the neuropathologists included in the present investigation evaluated subsets of the same 62 histological cases but did not examine cases provided from their own laboratories and were unaware of clinical information, the neuropathologists' accuracy measures were in general comparable to the neural network (Table 1). The supervised ANN performed better or similarly to the neuropathologists in classifying PSP, atypical PSP and PEP but had both lower specificity and predictive values for corticobasal degeneration and Pick's disease. Measures of accuracy, in particular positive predictive value,
Table 3  Unsupervised ANN

<table>
<thead>
<tr>
<th>Similarity threshold (0.5)</th>
<th>Similarity threshold (0.6)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group 1</strong> ($n = 30$)</td>
<td><strong>Group 1</strong> ($n = 20$)</td>
</tr>
<tr>
<td>Typical PSP</td>
<td>Typical PSP</td>
</tr>
<tr>
<td>Atypical PSP</td>
<td>Atypical PSP</td>
</tr>
<tr>
<td>Corticobasal degeneration</td>
<td>Corticobasal degeneration</td>
</tr>
<tr>
<td></td>
<td>Group 2 ($n = 32$)</td>
</tr>
<tr>
<td>Postencephalic parkinsonism</td>
<td>Postencephalic parkinsonism</td>
</tr>
<tr>
<td>Pick’s disease</td>
<td>Pick’s disease</td>
</tr>
<tr>
<td>Corticobasal degeneration</td>
<td>Corticobasal degeneration</td>
</tr>
<tr>
<td>Atypical PSP</td>
<td>Atypical PSP</td>
</tr>
<tr>
<td>Group 3 ($n = 9$)</td>
<td>Group 3 ($n = 9$)</td>
</tr>
<tr>
<td>Pick’s disease</td>
<td>Pick’s disease</td>
</tr>
<tr>
<td>Corticobasal degeneration</td>
<td>Corticobasal degeneration</td>
</tr>
<tr>
<td>Group 4 ($n = 8$)</td>
<td>Group 4 ($n = 8$)</td>
</tr>
<tr>
<td>Corticobasal degeneration</td>
<td>Corticobasal degeneration</td>
</tr>
<tr>
<td>Group 5 ($n = 1$)</td>
<td>Group 5 ($n = 1$)</td>
</tr>
<tr>
<td>Corticobasal degeneration</td>
<td>Corticobasal degeneration</td>
</tr>
<tr>
<td>Group 6 ($n = 1$)</td>
<td>Group 6 ($n = 1$)</td>
</tr>
<tr>
<td>Corticobasal degeneration</td>
<td>Corticobasal degeneration</td>
</tr>
</tbody>
</table>

Values are percent of cases in each group. *Atypical PSP cases overlapped between Groups 1 and 2; **corticobasal degeneration cases overlapped between groups 2, 4, 5 and 6.

may be heavily influenced by the relative frequency of the target disease. Although the relative frequency of the disorders studied was purposefully chosen to be low to closely reflect the rarity of each of these disorders, both studies had similar relative frequency except for atypical PSP which had a higher frequency in the present study (0.15 versus 0.08). Thus, both studies also demonstrate the still limited neuropathological characterization of the disorders studied.

Since there are no biological markers for the diagnosis of these neurodegenerative disorders, clinicopathological diagnosis is the proposed ‘gold standard’ for their diagnosis (Khachaturian, 1985; Mirra et al., 1991; Hauw et al., 1995; Litvan et al., 1996). In this regard, it would have been desirable, in some respects, to create an independent panel of expert neuropathologists to review the cases and by consensus arrive at the correct diagnoses. However, for budgetary and operational reasons, we chose to accept as correct the diagnosis of each case given by the neuropathologist who donated and reviewed the case, based on the NINDS criteria (Hauw et al., 1994). To increase the relevance of our findings, the neuropathologists who extracted the data values for input into the ANN (morphological features) were familiar with the classification of neurodegenerative disorders and were most of the same as the ones who had previously participated in a blind-study and achieved substantial inter-observer reliability when they evaluated the same diseases (Litvan et al., 1996). Thus, substantial inter-observer agreement of neuropathologists probably helped minimize differences in data extraction. In addition, since the addition of minimal clinical information improves the accuracy of the diagnosis (Litvan et al., 1996), we chose that the neuropathologists extracting the data for input into the ANN were aware of the clinical information.

Future studies designed to examine more frequent disorders may benefit from having only one expert neuropathologist evaluate the specimens. Furthermore, to enhance objectivity, in the future, perhaps, appropriate image analysis (Becker, 1994; Rogers et al., 1994; Pizzi et al., 1995; Dybowski and Gant, 1995) and feature detection algorithms could be developed as front-end processors to diagnostic classification ANNs.

The performance of the supervised ANN was relatively similar to that found in previous studies using ANN methodology in other fields of medicine (Reggia, 1993; Christy et al., 1995). However, it should be mentioned that it is difficult to compare the accuracy of the neuropathologists with the ANN. Input data to the ANN consisted of variables derived by neuropathologists examining their own histological specimens and not by the computer examining the slides directly. Neuropathologists may have used ‘hidden’ disease markers which were not provided to the ANN, such as intensity of staining of the neuropil or compactness of the neuropil as being the reflection of particular disease patterns. Similarly, neuropathologists may have benefitted from evaluating other histological features to categorize the chronicity of the disease such as the type of gliosis (i.e. active microglia in disorders of shorter duration versus astrocytes in those with of longer duration) or used as a diagnostic aid other features such as free pigment, unusual in PEP but more common in acute disorders. On the other
hand, the ANN may require extense or severe histopathological lesions rather than few histopathological markers to cluster disorders. Currently, there are no practical methods for the computer to directly obtain the raw data from the original histological samples. Since human experts are needed to extract the neuropathological data values for input to the ANN, a direct comparison of the experts and an ANN is not fair to either.

The other main use of the supervised ANN methodology in our study was to identify histological features that could help differentiate the studied disorders. In this regard, the methodology supported currently proposed criteria for PSP, corticobasal degeneration and Pick’s disease (Hauw et al., 1994; Litvan et al., 1996). It also suggested a few histological features that might help differentiate PEP from other studied neurodegenerative disorders that need to be prospectively evaluated for their significance. However, it did not detect features to characterize atypical PSP, reinforcing the notion that this may be an artificial entity (Litvan et al., 1996).

**Experiment 2**

The difficulties in differentiating neuropathologically some of the atypical parkinsonian disorders (Litvan et al., 1996), and the existence of overlapping cases (Seitelberger et al., 1983; Arima et al., 1992; Gibb, 1992; Daniel et al., 1994; Lennox et al., 1994; Jendroska et al., 1995) suggest the possibility that these disorders are not different nosological entities but rather a single disease with a range of neuropathological and clinical manifestations. However, the higher threshold (0.6) unsupervised learning ANN study suggests that these disorders cluster into four major groups, supporting the current nosological classification as different entities, with some overlapping cases that may ‘link’ these disorders. Typical PSP, Pick’s disease, PEP and less strongly corticobasal degeneration, clustered in different groups. However, atypical PSP crossed-over into both typical PSP and PEP classifications. These results coincide with the difficulty neuropathologists experience in separating atypical PSP cases from typical PSP and PEP cases (Litvan et al., 1996). They also point to the relative difficulty in separating some cases of corticobasal degeneration from PSP and Pick’s disease (Gibb, 1992; Lennox et al., 1994; Jendroska et al., 1995 Litvan et al., 1996). The results stress the heterogeneity of atypical PSP and of corticobasal degeneration that was present in our sample. These findings also point to the still hard task of differentiating corticobasal degeneration from these last two entities (Uchihara et al., 1994; Feany and Dickson, 1995), and the need to find features to distinguish them (Gibb, 1992). Cases clustered in group two seem to have less florid pathology or include borderline cases, again suggesting that the ANN may require extense lesion distribution or severe density to distinguish these disorders.

Interestingly, with the lower threshold (0.5), atypical PSP cases overlapped between two groups, and a few corticobasal degeneration cases were hard to separate from typical PSP. The fact that the 0.5 threshold unsupervised learning ANN mainly separated PSP from non-PSP cases suggests that PSP is a distinctive entity. On the other hand, borderline cases emphasize that the separation between these entities is relative.

The unsupervised learning ANN seems to be a method to explore the existence of alternative neuropathological classifications of neurodegenerative diseases. A possible limitation of this method as implemented in this study is that all covariates were given equal weights. Future studies may benefit from exploring the use of different weights according to the possible diagnostic relevance of the distinct histological markers identified.

**The way forward**

We chose to use the Dystal ANN paradigm because we anticipated that it would perform the diagnostic classification tasks well, as it did, and because the supervised and unsupervised learning processes in Dystal produce similar patches which can be studied and compared to gain causal insight to human and machine diagnostic classifications. Dystal is sufficiently robust to allow diverse clustering representing the same disease entity and yet contained enough to provide classification understanding through the interpretation of patches. This understanding or explanation feature is not as available in more robust ANN paradigms such as back error propagation. Furthermore, unlike other ANN training paradigms, Dystal ANN training can be additively upgraded as new training advances arrive. This means that it does not have to be retrained from the beginning like other ANN paradigms (e.g. back error propagation). It is expected that Dystal would emulate the performance of human experts whose explicit or implicit classification rules are based on predictive covariates. Dystal could start to falter, however, as covariate interactions become more complex, in which case more robust ANN paradigms like back error propagation could offer improved performance.

Artificial neural network methodology appears useful for exploring neuropathological classifications. We were able to use it to identify features that may help differentiate the particular neuropathological disorders studied as well as to evaluate alternative neuropathological classifications. Ideally, advanced technology using hardware or software image analysis and feature detection algorithms could be developed as front-end processors to diagnostic classification ANNs. Unsupervised learning ANN methods could be enhanced when hybridized with expert systems components that model higher level cognition of expert human diagnosticians. Such technology may then be able to provide further insights into the nosological classification and biology of these complex disorders. In general, neural network methodology seems highly promising for studying neurodegenerative disorders.
Acknowledgements
S.E.D. received support from a grant from the Parkinson's Disease Society of the United Kingdom, and P.L.L. was supported by the Medical Research Council.

References


Received December 1, 1995. Accepted January 12, 1996