Correlation between cognitive function scores and the response of a neural network classifier for SPECT data in patients with Alzheimer's disease

Sren Halkjr[1], Gunhild Waldemar[2], Benny Lautrup[1] and Olaf B Paulson[2]

1. CONNECT, Niels Bohr Institute, University of Copenhagen, Denmark
2. Department of Neurology, Rigshospitalet, University Hospital, Copenhagen, Denmark

Corresponding author and reprint requests:

Gunhild Waldemar

Department of Neurology, N2082, Rigshospitalet

University Hospital, Blegdamsvej 9

DK-2100 Copenhagen, Denmark

Fax: +45 35 45 26 26

Running footline: Neural Network Analysis of SPECT Images

Abstract

An artificial neural network (ANN) was used to analyse and classify [^{99m}Tc]-d,l-HMPAO SPECT data sets from 25 patients with a diagnosis of probable Alzheimer's disease and 25 healthy control subjects. **Methods:** Prior to the ANN analysis, the original set of 8 semiquantitative CBF values from cortical ROIs representing each subject was transformed into a new set of 8 values consisting of sums and numerical differences between corresponding left and right regions. An ANN was then trained on the binary classification problem based on these values. During training the ANN was *pruned* in order to discover the optimal ANN architecture. After training, a calculation of the *saliencies* of the ANN parameters ranked the lobar regions according to their importance in the classification task. The continuous output of the ANN was compared to the corresponding Mini-Mental State Examination (MMSE) scores in order to see whether the ANN output could be interpreted as a measure of disease severity albeit the ANN was trained on the binary classification task only. **Results:** The classification problem was found to be linearly separable. The resulting ROC curve of the linear ANN classifier had a ROC area equal to 0.93. A study of the *saliencies* of the ANN parameters showed that the response of the ANN was largely based on 4 out of the 8 input variables. The correlation coefficient between the response of the linear ANN classifier and the MMSE-scores was -0.7 ($p < 0.001$).

Conclusion: By using the method of pruning the classification task has been found to be linearly separable. The ROC area of the linear ANN classifier was comparable to that of more complex ANN classifiers found in similar analyses. The concept of parameter saliencies may provide important information about the involved regions. Although trained on the binary classification problem only, the significant correlation between MMSE scores and the corresponding continuous output of the linear ANN indicates that the ANN output provides information about the disease severity. Thus, the MMSE scores reflect in a natural (linear) way the changes in rCBF values.

Key Words: Alzheimer's disease, MMSE scores, ANNs, pruning, parameter saliencies, .

Introduction

Artificial neural networks (ANNs) [1, 2, 3] have in recent years proven to be a valuable supplement to conventional discriminant analysis in medical pattern recognition tasks. An example of such a task is the analysis and interpretation of Single Photon Emission Computed Tomography (SPECT) or Positron Emission Tomography (PET) brain images in Alzheimer's disease (AD). This task has been studied in a few reports [4, 5, 6, 7] where it was found that the ROC performance [8] of the ANN was superior to that of a linear or quadratic discriminant and comparable or even superior to that of a human expert. This success may partly be due to the *non-parametric* nature of ANN modeling: a functional form of the discriminant function (f. ex. linear or quadratic) is not assumed a priori, rather it is chosen by the task itself. This model selection may be improved by the method of *pruning* [9] which provides an important tool for choosing the right size and architecture of the ANN. Furthermore, out of a (large) number of input variables, this method can point out those important for the discrimination task. This may be extremely useful in an area such as brain modeling due to the (possibly) limited knowledge about regional importances. A number of preprocessing methods for increasing model convergence exist of which *subject* and *region centering* are among the most popular. In the studies of Kippenhan et al. [4, 5] observation centering has been observed to increase model convergence as well as improving the ROC performance. While the improvement of ROC performance probably is problem specific since information is thrown away, the increase in convergence due to observation centering can be justified theoretically [10, 11].

In this article we describe the application of an ANN to the classification and evaluation of SPECT images of normalized regional cerebral blood flow (rCBF) values in patients with probable AD and healthy controls. Patients with a clinical diagnosis of AD are characterized by having reduced rCBF as well as asymmetric rCBF deficits in the frontal, parietal and temporal association cortex while the occipital cortex, basal ganglia and cerebellum are relatively unaffected [12]. Based on this knowledge, 8 rCBF values were selected for the ANN analysis, namely the sum of and the numerical difference between the rCBF values of the left and right temporal, frontal and parietal cortex and hemisphere. By incorporating this information into the data set, we aimed to make the task simpler, thus requiring a smaller ANN. Such an ANN will be faster to train and easier to interpret. Although the ANN was trained only to separate the two patient groups, the response of the ANN for each patient was compared to the corresponding MMSE score [13] to investigate whether the ANN captured the severity of the disease by itself. This investigation is different from *training* the ANN to actually reproduce the MMSE scores [4].

Materials and methods

Patients

The study reanalyzed [^{99m}Tc]-d,l-HMPAO SPECT data sets from 25 patients with a diagnosis of probable Alzheimer's disease and 25 healthy control subjects, which were presented in detail in a previous publication [14]. Briefly, the patients all fulfilled the DSMIIIR criteria [15] for dementia and the clinical criteria from the *National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association* (NINCDS-ADRDA) [16] for probable Alzheimer's disease. Their mean age was 70 years (range 53-83), and their mean Mini-Mental State Examination (MMSE) [13] score was 15 (range 3–27). The 25 control subjects were carefully screened age-matched healthy, non-hospitalized volunteers, presented in detail in a previous publication [17]. Their median age was 70 years (range 53-83). Cranial CT was without focal pathology in all 50 patients and control subjects.

SPECT

A saline bolus containing 1.1 GBq [^{99m}Tc]-d,l-HMPAO (Exametazime, Ceretec, TM Amersham, London, United Kingdom) was injected intravenously 10-20 min before data acquisition. The activity in the brain was measured by the Tomomatic 64 (Medimatic, Hellerup, Denmark), a 3 slice brain dedicated SPECT instrument. With this technique the slice thickness, and the resolution in the plane is 10 mm, full width at half maximum (FWHM). All scans were obtained in parallel with the cantho-meatal plane. The acquisition time was 25 min., yielding at least 3×10^6 counts per slice. Data acquisition was performed at 3 different levels in order to obtain 9 contiguous image slices covering the whole brain [17]. Following reconstruction, the data were normalized to the cerebellum and corrected to adjust for incomplete retention of the tracer by the algorithm described by Lassen et al. [18], using a conversion/clearance ratio of 1.5. The cerebellar hemisphere with the highest countrate was used as the reference region, and its countrate was determined by a lower 60% fractile threshold [17]. The procedure for regional analysis of [^{99m}Tc]-d,l-HMPAO data was described in detail in a previous publication [17]. Briefly, regions of interest (ROIs) were drawn on each slice with reference to pre-defined standardized templates with bilateral and symmetric ROIs. The design of these templates was performed with reference to an anatomical atlas of the brain [19] and based on anatomical structures which are easily recognizable on a typical SPECT image [17]. For each ROI, i, the mean pixel value and the area of the region were determined in each slice in which the ROI appeared [17], and a weighted mean rCBF value was calculated for the entire left ($\mathrm{F}_{i(L)}$) and right ($\mathrm{F}_{i(R)}$) ROI, including all slices in which the ROI appeared. In this presentation 8 rCBF values were included: mean rCBF in the left and right frontal cortex ($\mathrm{F}_{f(L)}$ and $\mathrm{F}_{f(R)}$), mean rCBF in the left and right temporal cortex ($\mathrm{F}_{t(L)}$ and $\mathrm{F}_{t(R)}$), mean rCBF in the left and right parietal cortex ($\mathrm{F}_{p(L)}$ and $\mathrm{F}_{p(R)}$), and mean rCBF in the entire left and right hemisphere ($\mathrm{F}_{h(L)}$ and $\mathrm{F}_{h(R)}$).

ANN analysis

Prior to the ANN analysis, the original set of 8 rCBF values for each subject was transformed into a new set of eight values. For each pair of corresponding left and right values in the original set, a new pair consisting of the sum and the numerical difference between these was created e. g. $(F_{\cdot(L)}, F_{\cdot(R)}) \rightarrow (F_{\cdot(L)} + F_{\cdot(R)}, |F_{\cdot(L)} - F_{\cdot(R)}|)$. Thus, the new data set stressed the importance of global blood flow values and blood flow asymmetries in the classification task. Note that by taking the numerical difference, the transformation becomes non-linear.

In order to speed up learning [20], the transformed data set was *region centered.* Let x_{rs} denote the blood flow through region r in subject s and $p = 50$ the total number of subjects. The average blood flow in region r across subjects is $x_{r\cdot} = \sum_{i=1}^{p} x_{rs}/p$. This was subtracted from each subject giving

$$\tilde{x}_{rs} = x_{rs} - x_{\cdot s}$$

As the covariance matrix is invariant under this transformation, the classification problem is effectively unaltered.

We used a standard representation of a one-hidden-layer ANN [3] characterized by the number of hidden units and with a sigmoidal output in the range $[-1, 1]$. Patients with AD were labeled +1 while control subjects were labeled -1. $\mathbf{w}$ was used to denote the set of ANN parameters and as error function we used the entropic error

$$E[\mathbf{w}] = \sum_{j=1}^{p} \frac{1}{2}(1+t^j)\log\frac{1+t^j}{1+o^j} + \frac{1}{2}(1-t^j)\log\frac{1-t^j}{1-o^j}$$

where the dependence on the ANN parameters has been shown explicitly. Here t^j denote the patient labels described above while $o^j \in [-1, 1]$ denotes the ANN output when presented with input vector j. The convergence properties of this error function are superior to those of the squared error function [21]. The error function was then minimized with respect to the ANN parameters using the standard iterative gradient descent algorithm [3]. The training of the ANN was stopped once the sign of the ANN output was right for all subjects, since then it is only a matter of scaling the ANN parameters to achieve the desired output.

Prior to training the ANN, one must decide upon a particular architecture of the ANN (choose the number of hidden units). Then, to adapt the architecture of the ANN during training we used the method of pruning [9] in a simplified form. After the ANN had been trained to a minimum $\mathbf{w}^0$ where it was able to classify all subjects correctly, the *saliency* s_i of each ANN parameter w_i was estimated as

$$s_i \equiv E(w_i = 0) - E(w_i = w_i^0)$$

Thus the saliency of a parameter is a measure of how much the error function increases if that parameter is deleted (zeroed). Ideally, one would want to calculate the saliency using the error after the ANN had been retrained with the particular weight clamped to zero. However, experience shows that the two ways of ranking the parameters through their saliencies are identical. The parameter with the smallest saliency was then zeroed after which the ANN was retrained. As we defined the right ANN architecture as the smallest ANN capable of learning the data set, this was repeated until no more parameters could be zeroed without introducing misclassifications into the data set.

The performance of the smallest found ANN was estimated through the area under the ROC curve [8]. A ROC curve is generated by testing the ANN on new subjects using different decision thresholds. For each decision threshold two performance characteristics are obtained, namely the *false positive* ratio and the *true positive* ratio, which are plotted against each other. Using the area under this curve, a measure of the performance is obtained which is independent of a particular decision threshold. A leave-two-out cross-validation was used to test the ANN. The ANN was tested on a patient with AD and a control subject while being trained on the remaining examples. Twenty-five such pairs could be generated and the points on the ROC curve were averages over these. The decision thresholds were chosen in the ANN output range $[-1, 1]$ with an intermediate distance equal to 0.001. In a group of different points with identical x-axis values, the point with the largest y-axis value (best performance) was chosen. The area under the ROC curve was calculated using a standard cubic-spline-interpolation based method.

Results

Figure 1 shows the mean pattern vectors together with the standard deviations for the subject groups before being region centered. The rCBF values are clearly higher in the control group than in the Alzheimer group, while the asymmetries in blood flows are higher in the Alzheimer group than in the control group. Note that these characteristics were not spoiled by the *region centering* since we subtracted the mean pattern vector across the entire data set from each input vector.

In order to characterize the data set, an ANN with initially three hidden units was trained on the entire *region centered* data set. Since this ANN was able to classify all subjects correctly, ANNs with two and one hidden units respectively were successfully trained on the data. It is easy to realize that an ANN with one hidden unit is in fact a linear discriminant function

$$\Theta(\mathbf{x}_s) = \mathrm{sign}(\mathbf{w} \cdot \mathbf{x}_s - \theta)$$

This means that the two subject groups were linearly separable. The following analysis was therefore confined to the linear ANN equal to the weighted sum expression $\mathbf{w} \cdot \mathbf{x}_s - \theta$.

After a normalization of the weighted sums to the unit interval these were plotted against the corresponding MMSE scores together with a best-fit line in the squared error sense. A typical result is shown in figure 2. The intersection between the best-fit line and the horizontal line suggests a threshold equal to 22 on the MMSE scale for separating healthy control subjects from patients with AD. An average correlation between MMSE scores and weighted sums over twenty runs was equal to -0.7 with negligible standard deviations ($p < 0.001$).

Figure 3 shows the saliencies of the different input variables. However, these saliencies were calculated using the value of the error function after the ANN had been retrained in order to be more reliable (as discussed earlier). According to this plot, the diagnoses of the ANN were to a large extent based on the difference between the left and right hemisphere. However, it is seen that no input variable can be deleted without introducing misclassifications.

To estimate the performance of the ANN classifier the average area of twenty ROC curves over twenty runs was calculated ($A = 0.93 \pm 0.002$). A typical example of a ROC curve is shown in figure 4. Optimal performance for a single decision threshold was $87.6\% \pm 0.8\%$ correct classification which was obtained for decision thresholds in a small interval around zero.

Finally, table 1 shows the signs and factors of the parameters in the ANN together with their corresponding input variables.

Discussion

This report demonstrates the ability of an ANN classifier to separate SPECT image sets with significant hypoperfusion in patients with a clinical diagnosis of AD from image sets obtained in healthy controls.

Conventional methods for image analysis comprise various methods for ratings based on subjective visual analysis and various ROI based parametric image analysis methods all associated with certain limitations. Visual rating methods are heavily dependent on observer experience and associated with possible observer bias. ROI based methods often do not include information from the entire image sets but only from selected ROIs. The definition of these ROIs may be associated with some observer bias. The statistical significance of small changes in one ROI when a larger number of ROIs are analyzed in the same patient is not clear. Arbitrarily changes in internal ratios exceeding 2 standard deviations have been taken as a significant change.

Using a neural network classifier for the evaluation of a SPECT image set has the advantage of being a non-observer dependent, quick and easy method for identifying an image set as abnormal or normal. The diagnostic accuracy of the ANN analysis may not be better than that of visual or ROI based analysis alone but has the advantage of being non-observer dependent. The result will not depend on the experience or bias of any observer, the method can be used across different centers and for different imaging methods, and the results are consistent from one analysis to the other. The most important limitation of the technique is that regional information is not required and must be obtained by conventional methods.

Regarding the particular data used here, by pruning the ANN it was found that the classification problem was linearly separable. However, the size of the data set is small compared to the sizes of the data sets analyzed in similar articles [4, 5, 6, 7] and it is an open question whether this will hold true for larger data sets. However, it is our belief that the size of the found ANN is closer to the "true" model than the (much) more complex ANNs found in the above articles. This is supported by the ROC area which is comparable to and in some cases superior to those found in these articles. Furthermore, the fact that the average MMSE score of the patient group is comparable to those in the above mentioned articles indicates that this classification problem was as difficult as those problems.

Due to linear separability of the transformed data set, it is concluded that the non-linear transformation prior to the ANN analysis served it's purpose, since the original data set was not linearly separable. This shows that the data analysis may be simplified by incorporating available human knowledge.

The saliencies show that none of the input variables can be deleted without introducing misclassifications. This shows, that when the analysis is based on a set of large regions, the union of these regions should cover the whole brain. More detailed information could be gained if the analysis was based on a (large) number of smaller regions. In this case, pruning could rule out the unimportant regions, so that further analysis could be more focused. This was observed in [11] where optimal ANN performance in a similar classification problem was achieved using a set of 23 ROIs out of an original set of 40 ROIs.

Although the ANN was trained only to solve the binary classification task, it was found that the continuous output of the ANN correlated reasonably well with the corresponding MMSE scores. It is interesting to note that the ANN finds it's own MMSE threshold equal to 22 (see figure 3) for separating control subjects from patients with probable AD. This is close to the threshold found in [13] equal to 20. This observation indicates that the MMSE scale in a natural (linear) way reflects the changes in rCBF during progression of the disease.

Conclusion

We have shown how the use of available human knowledge together with the method of pruning have reduced the complexity of a classification task of SPECT images using an ANN. The result is a simple ANN which is fast to train and easy to interpret. Furthermore, pruning provides information about the importance of the involved variables (ROIs). However, regarding the latter, the effect of pruning will be more pronounced in a data set with a larger number of small regions than here.

Our analysis indicate that there may be a natural (linear) correspondence between the MMSE scale and changes in rCBFs.

We suggest that classification of image sets using pruned ANNs could be included in the routine analysis of SPECT as an initial evaluation supplemented by visual or parametrial regional analysis of image sets classified as abnormal. For future multi center drug trials, SPECT may be included as one of many effect parameters, and here an unbiased non-observer dependent and consistent image analysis method is necessary. Further studies in which the ANN has been trained on larger patient samples, on more than 2 groups (classes) of patients and on milder cases are needed.

References

[1] Brian D. Ripley. *Pattern Recognition and Neural Networks.* Cambridge University Press, 1996.

[2] Christopher M. Bishop. *Neural Networks for Pattern Recognition.* Oxford University Press, 1995.

[3] John Hertz, Anders Krogh, and Richard G. Palmer. *Introduction to the theory of neural computation.* Redwood City, CA: Addison-Wesley, 1991.

[4] J. S. Kippenhan, W. W. Barker, J. Nagel, C. grady, and R.Duara. Neural-Network Classification of Normal and Alzheimer's Disease Using High-Resolution and Low-Resolution PET Cameras. *J Nucl Med*, 35:7–15, 1994.

[5] J. Shane Kippenhan, Warren W. Barker, Shlomo Pascal, Joachim Nagel, and Ranjan Duara. Evaluation of a Neural-Network Classifier for PET Scans of Normal and Alzheimer's Disease Subjects. *J Nucl Med*, 33:1459–1467, 1992.

[6] Michael P. A. Page, Robert J. Howard, John T. O'Brien, Muriel S. Buxton-Thomas, and Alan D. Pickering. Use of Neural Networks in Brain SPECT to Diagnose Alzheimer's Disease. *J Nucl Med*, 37:195–200, 1996.

[7] Karen H. Chan, Keith A. Johnson, J. Alex Becker, Andrew Satlin, Jack Mendelson, Basem Garada, and B. Leonard Holman. A Neural Network Classifier for Cerebral Perfusion Imaging. *J Nucl Med*, 35:771–774, 1994.

[8] Charles E. Metz. Roc methodology in radiologic imaging. *Invest Radiol*, 21:720–733, 1986.

[9] Yann Le Cun, John S. Denker, and Sara A. Solla. Optimal brain damage. *Advances in Neural Information Processing Systems*, 2:598–605, 1990.

[10] Søren Halkjær and Ole Winther. The effect of correlated input data on the dynamics of learning. *Advances in Neural Information Processing Systems*, 1997.

[11] Søren Halkjær. Dynamics of learning: application to the diagnosis of HIV and Alzheimer patients. Master's thesis, University of Copenhagen, April 1996.

[12] Gunhild Waldemar. Functional Brain Imaging with SPECT in Normal Aging and Dementia. *Cerebrovasc Brain Metab Rev*, 7(2):89–130, 1995.

[13] MF Folstein, SE Folstein, and PR McHugh. "Mini-Mental state": a practical method for grading the mental state of patients for the clinician. *J Psychiatr Res*, 12:189–198, 1975.

[14] G Waldemar, P Bruhn, M Kristensen, A Johnsen, OB Paulson, and NA Lassen. Heterogeneity of neocortical cerebral blood flow deficits in dementia of the Alzheimer type: ^{99m}Tc-d,l-HMPAO SPECT study. *J Neurol Neurosurg Psychiatry*, 57:285–295, 1994.

[15] American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*. Washington, D.C.: American Psychiatric Association, 3rd edition - revised. DSM-III-R edition, 1987.

[16] G McKhann, D Drachman, M Folstein, R Katzman, D Prize, and EM Stallan. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRA Work Group under the auspices of Department of Health and Human Services Task Forces on Alzheimer's Disease. *Neurology*, 34:939–944, 1984.

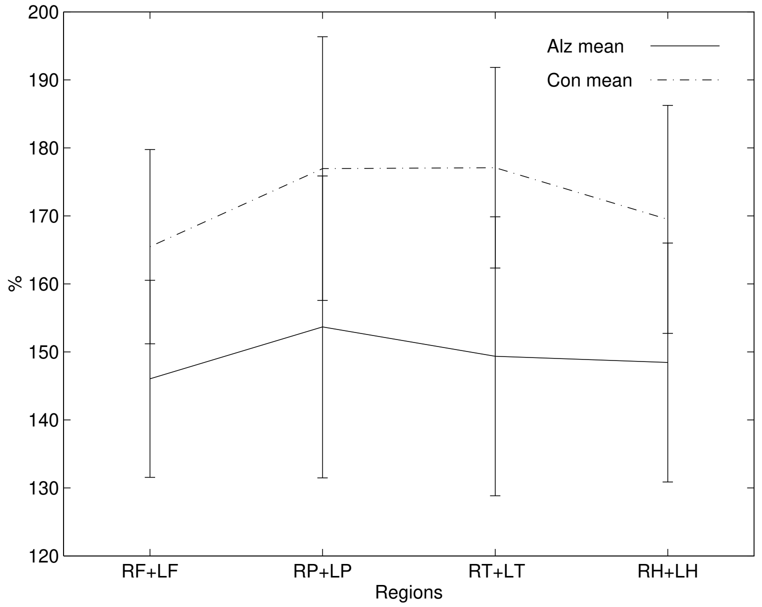
[17] G. Waldemar, S. G. Hasselbalch, A. R. Andersen, F. Delecluse, P. Petersen, A. Johnsen, and O. B. Paulson. ^{99m}Tc-d,l-HMPAO and SPECT of the Brain in Normal Aging. *J Cereb Blood Flow Metab*, 11:508–521, 1991.

[18] N. A. Lassen, A. R. Andersen, L. Friberg, and O. B. Paulson. The Retention of [^{99m}Tc]-d,l-HMPAO in the Human Brain after Intracarotid Bolus Injection: A Kinetic Analysis. *J Cereb Blood Flow Metab*, 8:S13–S22, 1988.

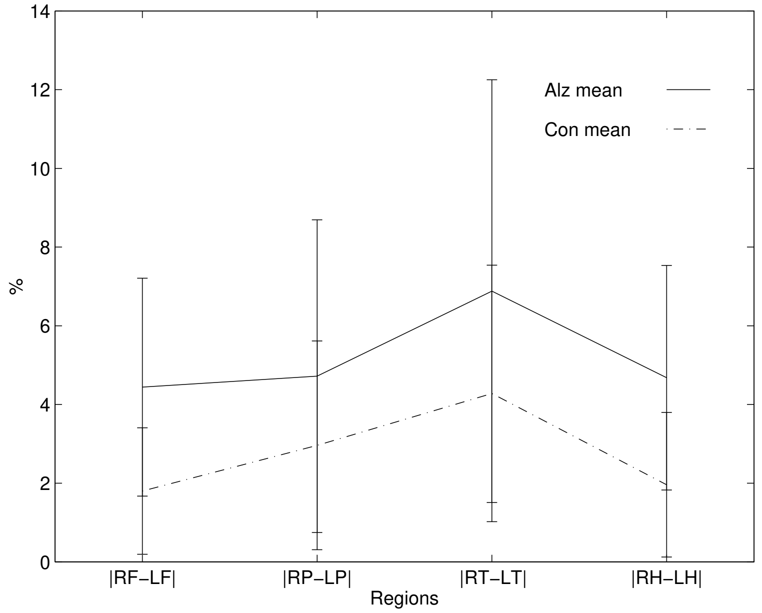
[19] S-M Aquilonius and S Eckernäs. *A colour atlas of the human brain. Adapted to computed tomography.* Stockholm, Esselte Studium, 1980.

[20] Yann Le Cun, Ido Kanter, and Sara Solla. Eigenvalues of covariance matrices: Application to neural-network learning. *Phys Rev Lett*, 66(18):2396–2399, 1991.

[21] Sara A. Solla, Esther Levin, and Michael Fleisher. Accelerated learning in layered neural networks. *Complex Systems*, 2(6):625–640, December 1988.


Alz mean
Con mean
200
190
180
170
160
150
140
130
120
%
RF+LF
RP+LP
RT+LT
RH+LH
Regions

(a)


Alz mean
Con mean
14
12
10
8
6
4
2
0
%
|RF−LF|
|RP−LP|
|RT−LT|
|RH−LH|
Regions

(b)

FIGURE 1. Mean values over the two subject groups for the four variables representing sums of rCBF values (a) and for the four variables representing numerical differences in rCBF values (b). RF/LF: right/left frontal cortex, RP/LP: right/left parietal cortex, RT/LT: right/left temporal cortex, RH/LH: right/left hemisphere.

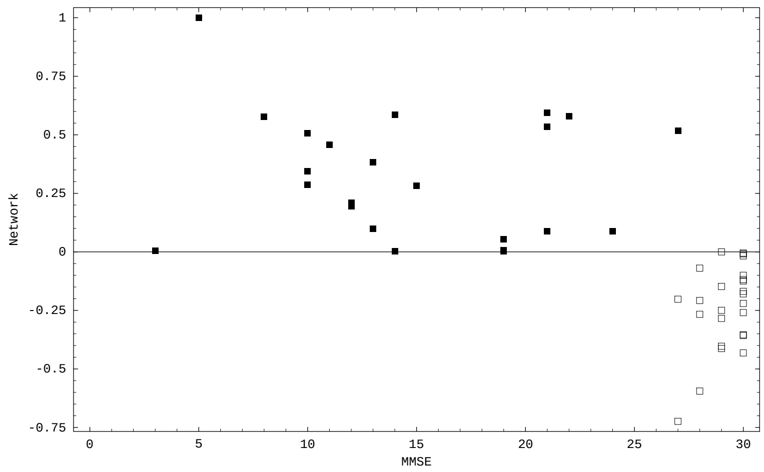


FIGURE 2. Scaled linear network output plotted against corresponding MMSE-scores for each patient. The horizontal line separates the two subject groups.

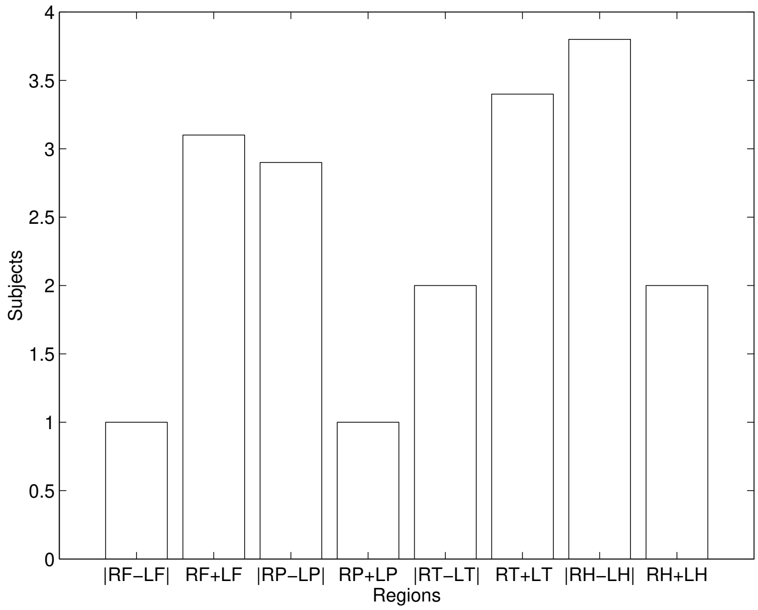


FIGURE 3. Saliencies of the eight input variables. These represent the number of misclassifications (on average) made by the network if that particular variable was left out.

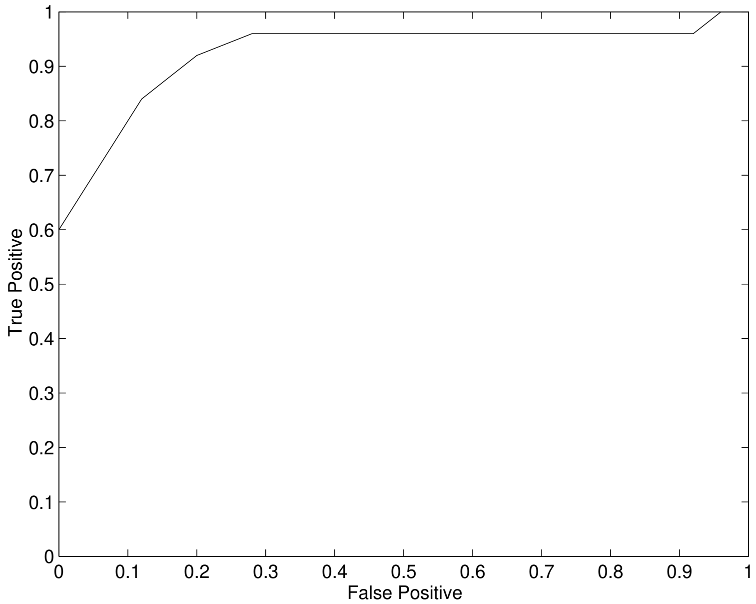


FIGURE 4. A typical ROC curve illustrating the classification performance of the linear neural network.

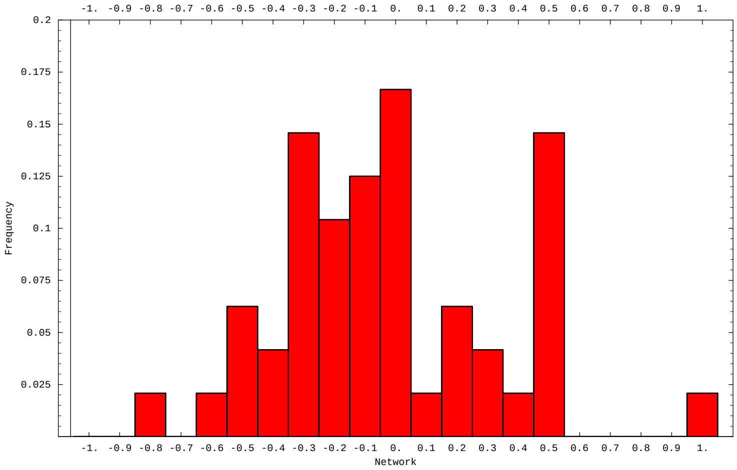


FIGURE 4. A typical ROC curve illustrating the classification performance of the linear neural network.

Table 1.
The values of the 8 parameters in the found ANN together with their corresponding input variables.

|RF-LF|	RF+LF	|RP-LP|	RP+LP	|RT-LT|	RT+LT	|RH-LH|	RH+LH
+1.5	-1.1	-1.4	-0.2	-2.2	+1	+5.2	1.9