



Published in final edited form as:

Neuroinformatics. 2004 ; 2(4): 399–416.

Artificial Neural Networks and Artificial Organisms Can Predict Alzheimer Pathology in Individual Patients Only on The Basis of Cognitive and Functional Status

Massimo Buscema^{*,1}, Enzo Grossi², David Snowdon³, Piero Antuono⁴, Marco Intraligi¹, Guido Maurelli¹, and Rita Savaré²

¹Semeion Research Center, Rome, Italy

²Bracco Medical Affairs Europe, Via E. Folli 50, 20134 Milan, Italy

³Sanders-Brown Center on Aging, Department of Preventive Medicine, College of Medicine, University of Kentucky, Lexington, KY

⁴The Medical College of Wisconsin, 9200 West Wisconsin Avenue, Milwaukee, WI

Abstract

Data from several studies have pointed out the existence of a strong correlation between Alzheimer's disease (AD) neuropathology and cognitive state. However, because of their highly complex and nonlinear relationship, it has been difficult to develop a predictive model for individual patient classification through traditional statistical approaches. When exposed to complex data sets, artificial neural networks (ANNs) can recognize patterns, learn the relationship of different variables, and address classification tasks. To predict the results of postmortem brain examinations, we applied ANNs to the Nun Study data set, a longitudinal epidemiological study, which includes annual cognitive and functional evaluation. One hundred seventeen subjects from the study participated in this analysis. We determined how demographic data and the cognitive and functional variables of each subject during the last year of her life could predict the presence of brain pathology expressed as Braak stages, neurofibrillary tangles (NFTs) and neuritic plaques (NPs) count in the neocortex and hippocampus, and brain atrophy. The result of this analysis was then compared with traditional statistical models. ANNs proved to be better predictors than Linear Discriminant Analysis in all experimentations (+ ~10% in overall accuracy), especially when assembled in Artificial Organisms (+ ~20% in overall accuracy). Demographic, cognitive, and clinical variables were better predictors of tangles count in the neocortex and in the hippocampus when compared to NPs count. These findings strengthen the hypothesis that neurofibrillary pathology may represent the major anatomic substrate of the cognitive impairment found in AD.

Introduction

Neurofibrillary tangles (NFTs) and neuritic plaques (NPs) are the primary neuropathologic markers of Alzheimer's disease (AD), although they can be found also in normal brain aging (Arriagada et al., 1992b; Giannakopoulos et al., 1995; Price and Morris, 1999; Giannakopoulos et al., 2003).

Neuropathological studies (Braak and Braak, 1991; Price and Morris, 1999; Giannakopoulos et al., 2003) have shown that the distribution of NFT in the human brain follows a predictable and hierarchical pattern, whereas the distribution of plaques varies among individuals. Neurofibrillary pathology is initially limited to the hippocampus and the entorhinal cortex

* Author to whom all correspondence and reprint requests should be addressed. E-mail:enzo.grossi@bracco.com

(Braak and Braak, 1991). As the number of tangles increases in these areas, neurofibrillary pathology extends into the temporal cortex. Finally, tangles emerge and spread to the neocortical areas of the brain.

The literature has consistently pointed out the existence of a strong correlation between AD neuropathology and cognitive function (Morris et al., 1989; Braak and Braak, 1991; Price et al., 1991; Snowden et al., 1996; Petersen, 2000; Riley et al., 2002; Guillozet et al., 2003). The neuropathological and cognitive changes preceding AD begin decades before diagnosis is made (Braak and Braak, 1991; Price et al., 1991; Petersen, 2000; Riley et al., 2002). This suggests the possibility that neuropathology may affect cognitive ability and daily function long before the symptoms make the diagnosis of dementia clinically apparent.

Recent studies further suggest that NFTs have a closer relationship to cognitive function with respect to amyloid plaques, not only in AD but also in normal aging and mild cognitive impairment (Morris et al., 1989; Arriagada et al., 1992a; Guillozet et al., 2003). In particular, the presence of high NFT density in the entorhinal and hippocampus neurons is strongly correlated to memory impairment in normal aging, whereas NFT formation in neocortical areas is associated with clinically overt AD (Braak and Braak, 1991; Price and Morris, 1999; Giannakopoulos et al., 2003).

Nevertheless, the variability of neuropathologic findings and their complex relationships to cognitive status observed in aging and AD (Braak and Braak, 1991; Price et al., 1991; Giannakopoulos et al., 1995; Riley et al., 2002) limits the development of a predictive model for individual patient classification using traditional statistical methods.

These methods, based on a generalized linear model, have limited value in predicting outcomes, because of multiple interacting factors that affect the final outcome.

Large-scale epidemiological studies have attempted to provide algorithms to assess an individual's risk for developing AD on the basis of identified risk factors; however, results have been disappointing. This is because the predisposing factors can act simultaneously with complex and nonlinear interactions in the same patient.

The development of artificial neural network (ANN) computer algorithms was inspired by the highly interactive processing functions of the human brain. Like the brain, ANNs can recognize complex patterns, manage data, learn the hidden relationship among different variables, and resolve classification problems.

The purpose of the study was to evaluate the performance of different types of ANNs in predicting the number of AD lesions based only on cognitive and functional parameters.

The application of ANNs—viewed as a new approach to the study of complex clinical data—may aid in the development of strategies for early, noninvasive detection of AD pathology.

Further, the development of a reliable in vivo model to determine the severity of neuropathologic lesions may be helpful in determining the efficacy of new pharmacological interventions intended to delay the onset of dementia.

The Variables

Clinical Variables

One hundred seventeen subjects participated in this analysis. They were participants in the Nun Study (Snowdon et al., 1996; Riley et al., 2002), a longitudinal study on normal aging and AD. All subjects are catholic nuns in the “School Sisters of Notre Dame” congregation, living in

seven regions of the United States. Therefore, the sample was composed only of women, who were subjected to a series of cognitive and functional exams every year. All participants agreed to donate their brain for a postmortem examination.

At the time of their last cognitive exam, the participants were between ages 76 and 102 (mean age 87.77, standard deviation [SD] 6.14). At the time of death, participants were between ages 76 and 103 (mean 88.72, SD 6.03). The last cognitive test, therefore, was carried out just prior to the date of death. The educational level of the participants was rather high: 14 women had 8 yr of schooling, 9 had 12, 54 had 16, and 40 had 18 yr of schooling.

The 16 variables included are reported in Table 1.

The cognitive testing battery used in the Nun Study (Snowdon et al., 1996; Riley et al., 2002) was the “Consortium to Establish a Registry for Alzheimer's Disease” (CERAD) (Morris et al., 1989), which evaluates the capacity of memory, language, video-spatial ability, concentration, and orientation.

All the data obtained from the 16 variables were updated every year until the time of the participants' death.

Pathological Variables

The following variables were measured to determine the presence of AD pathology: number of NFT and NP per mm² in the hippocampus and neocortex, the degree of cortical atrophy, and the severity of Alzheimer pathology using the staging method developed by Braak and Braak (1991). The above assessments were carried out in the middle frontal gyrus (Brodmann area 9), inferior parietal lobule (areas 39/40), middle temporal gyrus (area 21), and the CA1 and subiculum of the hippocampus, using the modified Bielschowsky stain.

Methods

Artificial Neural Networks

ANNs are adaptive models for the analysis of complex data sets inspired by the analytical processes of the human brain. These systems are able to modify their internal structure and process according to a defined objective function. This functionality is the basis for the ANNs' learning capability. They are particularly suited for solving nonlinear problems.

In Supervised Networks the desired output (target) is defined, for each input vector, prior to the start of the learning process. The output is decided by the environment outside the network, not deduced from the input vector or from a part of it. These types of networks calculate an error function that measures the distance between the target as desired fixed output and their own output, and adjust the connection strengths during the training process to minimize the result of the error function (Buscema, 1998a; Buscema, 2002).

As we will see, the experiments carried out are based on:

- Three types of learning laws:
 - Bp: Back Propagation (standard) (Bridle, 1989; Buscema, 1998b; Price et al., 1991).
 - Sn: Sine Net (Semeion) (Buscema et al., 2000; Vomveg et al., 2003).
 - Bm: Bi-Modal Networks (Semeion) (Vomveg et al., 2003).
- Three types of ANNs topology:

Feed Forward (FF) (standard) (Buscema and Sacco, 2000).

Self-Recurrent ANN (SF), (Semeion) (Buscema, 1998b), Static (SA) and Dynamic (DA).

Temporal Associative Subjective Memory, (TASM) (Semeion) (Buscema, 1999a; Mecocci et al., 2002; Vomveg et al., 2003), Static (SA) and Dynamic (DA).

- Artificial Organisms, as complex combinations of networks:
 - T&T (Training and Testing, Semeion) (Buscema, 2001a; Buscema, 2001–2002; Vomveg et al., 2003).
 - I.S. (Input Selection, Semeion) (Buscema, 2001b; Vomveg et al., 2003).
- Supervised software (Semeion) (Buscema, 1999–2003), which allows the combination of each learning law with each type of topology, was used for all the experimentations.

The Validation Protocol

The validation protocol is a fundamental procedure to verify the models' ability to generalize the results obtained in the testing phase. In this study, we selected the protocol with the greatest generalization ability on data unknown to the model.

The procedural steps in developing the validation protocol are:

Subdivision of the data set randomly into two subsamples: the first called Training Set, and the second, called Testing Set.

Selection of a fixed ANN (and/or Organism), which is trained on the Training Set. In this phase, the ANN learns to associate the input variables with those that are indicated as targets.

Saving of the weights matrix produced by the ANNs at the end of the training phase, and all of the parameters used for the training.

Input of the Testing Set into the ANN, so that in each case, the ANN can express an evaluation based on the training just performed. This procedure takes place for each input vector, but every result (output vector) is not communicated to the ANN. In this way, the ANN is evaluated only in reference to the generalization ability that it has acquired during the Training phase.

Reliability Protocol

This general training plan was further developed to increase the level of reliability of the generalization of the processing models. We have used the 5×2 cross-validation test described by Thomas Dietterich (1998). The author shows how this protocol can be slightly more powerful than other statistical tests for algorithms that can be executed 10 times.

The protocol includes two procedures whose results have been compared: a first procedure of random generation of the training samples, and a second one to optimize the samples and variables.

In the first procedure (Fig. 1A), the sample of 117 subjects was randomly subdivided five times into two balanced subsamples: one for the training phase (Training) and one for the prediction (Testing). Ten experiments were performed by inverting the pairs of samples in such a way that the subsample used for testing is used for training and vice versa. This procedure allowed

us to obtain a reliable assessment of the error for all of the techniques employed. The Feed Forward ANNs (Bp, Sn, and Bm) are compared with two models of linear statistics: the linear discriminant analysis (LDA) applied to classification problems, and the linear regression (LR) applied to assessment of values problems.

The second procedure (Fig. 1B) included a process of sample optimization by the systems called training and testing and input selection. In detail, this phase includes three distinct steps.

Step 1: The T&T and I.S. systems were applied to optimize the training and testing samples, and to reduce the number of input variables.

Step 2: Feed Forward ANNs based on different laws of learning (Bp, Sn, and Bm) were applied to optimized samples. For each law of learning, 10 experiments were normally carried out, with different weights initialization.

Step 3: Artificial Organisms were applied to optimized samples based on the law of learning that obtained the best result in the second step. For each Artificial Organism, 10 elaborations were normally carried out.

Artificial Organisms

The T&T system (Buscema, 2001a; Buscema, 2001–2002; Vomveg et al., 2003) is a population of ANNs based on the evolutionary Genetic Doping Algorithm (GenD) (Buscema, 2004) developed by the Semeion Research Center. By means of iterative procedures of data preprocessing, T&T aims to an optimal distribution of the complete data set in two training and testing subsets. GenD has demonstrated greater effectiveness in terms of its speed and capacity of convergence on a broad range of problems and in resolving complex optimization problems with respect to the limits of the classic genetic algorithms. T&T includes two main phases: the preliminary phase, to evaluate the parameters of the fitness function to be used on the given data set and the computational phase, used to extract the best training and testing sets that partition the data set records.

The preliminary phase includes the configuration of a standard Bp inducer, whose parameters (number of hidden units and layers) and weights initialization are fixed. Several training trials using the same total set allow one to stabilize the configuration suitable to the available data set, as well as to establish a mean number of epochs, which is necessary to reach the total convergence. The choice of this simple type of neural network, instead of more complex ones, is mainly because of its convergence speed and its ability to address polarization issues correctly.

In the computational phase, the initial population of possible solutions evolves following the GenD algorithm. In order to reach the desired evolution in the minimal time, only the simplest options are used: a single tribe and global genetic operators (Crossover and Mutation).

The optimization algorithm distributes the original sample in two or more subsamples to obtain the maximum performance possible for ANN that is trained on the first sample and validated on the second. The score reached by each ANN represents its fitness, and its probability of evolution. It is possible to limit eventual optimistic polarizations in the evaluation of the performances, to invert the two samples and to consider the average between the two obtained estimates as fitness of the algorithm and an estimate of the quality of the model.

The substantial difference between this approach and a traditional one, which is based on the random subdivision of the original sample, is the possibility of taking advantage of all the information present in the data set.

If T&T system is able to use the different records of data sets to train and test inducers not on a chance basis, the I.S. system is able to weight the different variables of the available data set intelligently.

Input Selection system based on GenD, improves the optimization process, as a selection mechanism of the input variables of a fixed data set. As data collection is done by trying to include all variables that can have a connection with the event being studied, and as the relationships between the variables detected and the function of the process under exam could be not known, we have a series of variables that do not contain any real information regarding the process being examined. When inserted into the model, these variables cause an increase of noise and greater difficulty for the ANNs to learn data correctly.

In the nonlinear field, an instrument like the correlation does not exist. Consequently, a heuristic approach, which neglects other types of measures and concentrates on a set of inputs that provides the best performances in an ANN training model was chosen.

Given their nature, the T&T and I.S. systems can be viewed as data preprocessing organisms with a high capacity to single out the information useful to optimize the relationship between the input and output variables during the calculation process. To obtain this result, ANN populations which evolve, according to the criteria of the GenD evolutionary algorithm, are used to reach the best representation of variables and of sample size of the problem to be resolved.

Optimization of Samples

The last step of the optimization phase includes the training of the advanced neural networks on the basis of the training and testing set generated by I.S.

The predictive model, based on the ANNs, used to determine the value of ANNs in predicting the presence of pathological lesions in the brain from the analysis of clinical variables is compared with the two traditional statistic models: LDA and LR.

Results

We have carried out the experiments concerning the following six research targets.

Braak stage. Classification of severity of Alzheimer pathology subdivided into two classes

NFTs in the cortex

- Classification of number of NFT per mm^a into two classes
- Prediction of the number

NFTs in the hippocampus

- Classification of number of NFT per mm^a into two classes
- Prediction of the number

NPs in the neocortex

- Classification of number of NP per mm^a into two classes
- Prediction of the number

NPs in the hippocampus

- Classification of number of NP per mm^a into two classes.

- Prediction of the number

Brain Atrophy in the neocortex. Classification of the degree of atrophy into two classes

Before moving on to the analysis of the results produced by the different calculation systems, it is opportune to analyze the values of the linear correlation (R^2) which each of the 16 variables has with its respective targets.

We show the values in Table 2. We can see that the degree of linear correlation between the variables and the target has a tendency to be low. The variables' degree of nonlinearity with respect to the target appears high; therefore, the relation which ties the input to the output could be complex.

The results of the six experiments are summarized in Table 3.

Braak Stage

The level of severity of the neurofibrillary Alzheimer pathology was determined using the Braak and Braak method (Braak and Braak, 1991). The seven stages (from 0 to 6) of the Braak scale were grouped into two classes:

Stages 0–2, 58 patients, equal to 49.57% of the sample.

Stages 3–6, 59 patients, equal to 50.43% of the sample.

The two classes make up the target of the calculation models: each patient was classified as belonging to one of the two classes by the different calculation models based on the 16 input variables described.

Each calculation was carried out following the validation protocol described. In the random calculation five pairs of samples (training and testing), on which the systems were trained, were selected.

Three blocks of results were singled out in this experiment in terms of degree of provisional capacity: those obtained from the Organisms with Bp law of learning using the T&T and I.S. optimization system (89.92%), those obtained with the random protocol from the simple ANNs with SineNet (FF-Sn) law of learning (79.86%), and those obtained from the LDA (72.36%).

In the sample optimization phase, the I.S. system chose 8 of the 16 input variables. Details are illustrated in Table 4, column A.

Only these eight variables were used to train the samples of the Organisms. It is interesting to note that the I.S. system selected both variables with an inexistent or very low degree of linear correlation (Age $R^2 = 0.083$, Dress 0.050, Stand $R^2 = 0.027$, Eat–Drink $R^2 = 0.072$, and ApoE4 $R^2 = 0.068$), and variables with a slightly significant degree of linear correlation (WRCL $R^2 = 0.382$, VRBF $R^2 = 0.246$, and MMSE $R^2 = 0.308$).

Neurofibrillary Tangles in Neocortex

The numbers of NFT per mm² microscopic field were determined from Bielschowsky stained sections. NFT were counted in five microscopic fields with the highest numbers of tangles in the middle frontal gyrus (Brodmann area 9), inferior parietal lobule (areas 39/40), and middle temporal gyrus (Area 21).

We subdivided the data set into two classes on the basis of preliminary mathematical nonlinear processing, which established the cut-off to subdivide the sample. This approach was used also in three following experiments.

The first class included all of the patients who had an average value of neocortical NFT per $\text{mm}^2 < 1.0$ (47 cases); a second class included all of the patients who had an average value of NFT per $\text{mm}^2 < 1.0$ (70 cases).

Based on the input variables, different calculation models classified each patient as belonging to one of the two classes.

In this experiment, the T&T and I.S. systems were used even for the simple ANNs. This allowed a new set of calculations, which permits further comparisons. Results are expressed as mean values on 10 calculations.

Four levels of data have resulted from this experiment: statistic systems (LDA) with random samples obtained a mean accuracy equal to 73.80%; the best ANNs with random samples (FF-Sn) obtained a mean accuracy equal to 84.16%; the best ANNs with sample optimization systems (FF-Bp) obtained a mean accuracy equal to 87.83% and finally the SelfDABp (Self-Recurrent Dynamic Back Propagation) Organism with sample optimization systems reached the best result with a mean accuracy equal to 90.36%.

The variables chosen by the I.S. system were, in this case, 8 out of 16 (see Table 4, column B). Four of the variables belong to the cognitive and language functions section (WRCL $R^2 = 0.253$, BOST $R^2 = 0.321$, VRBF $R^2 = 0.287$, and MMSE $R^2 = 0.380$) with a significant degree of linear correlation, the other four have a nonsignificant linear correlation.

Neurofibrillary Tangles in Hippocampus

The numbers of NFT per mm^2 microscopic field were determined from Bielschowsky stained sections. NFT were counted in the five microscopic fields with the highest numbers of tangles in the CA1 and subiculum regions of the hippocampus.

In this testing, the data set was also divided into two classes: the first class included patients who had an average value of hippocampal NFT per $\text{mm}^2 > 10$ (68 cases), a second class included patients who had an average value of NFT per $\text{mm}^2 < 10$ (49 cases).

In the random protocol, the mean accuracy obtained from the LDA was equal to 75.52%. The ANN which obtained the best result in the random phase was the FF-Sn with an accuracy of 83.95%.

Of the three ANNs which were used with the T&T and I.S. optimization protocols, the best result was obtained by the FF-Bm with an accuracy of 86.68%.

The TasmSABm (Temporal Associative Subjective Memory Static Bi-Modal) obtained the best result, achieving an accuracy of 87.83%. The variables selected by the I.S. system were 6 out of 16 (see Table 4, column C).

The number of variables selected is lower, but equally distributed in the data set's four macrocategories: Demographic (Age death), ADL (Walk, Eat–Drink), cognitive and linguistic functions (WRCL, VRBF), and genetic (ApoE4). Even in this case, variables with low linear correlation together with others having slightly significant correlation were selected.

Neuritic Plaques in Neocortex

Maximum NP count was measured in the neocortex (maximum count per mm² field seen in frontal, temporal, and parietal lobes of the neocortex).

The sample of 117 cases was subdivided into two classes: the first class included patients with a maximum number of NP in the neocortex >1.5 (87 cases); the second class included patients with a maximum number of NP in the cortex <1.5 (30 cases).

In the random protocol, the mean accuracy obtained by the LDA was 69.77%.

The ANN which obtained the best result in the random phase was the FF-Sn with an accuracy of 83.68%.

In this case, only one ANN with T&T and I.S. optimization protocol, the FF-Bp, was used. The result obtained was of 88.93% accuracy.

The law of learning selected to carry out the calculations with the Organisms was the Bp. Among the four tests used, the SelfDABp obtained the best result with an accuracy of 89.92%.

I.S. selected 13 out of 16 variables (see Table 4, column D).

The selection carried out by the I.S. system was numerically more consistent and the choice was distributed on all the categories of variables of the data set. Excluded variables were, a. Age Death $R^2 = 0.028$ (among the demographic variables), b. Eat-Drink $R^2 = 0.055$ (among the variables relative to activities of daily life), and c. BOST $R^2 = 0.224$ (among the cognitive and linguistic variables).

Classification of Neuritic Plaques in Hippocampus

Maximum NP count was measured in the hippocampus (maximum count per mm² field seen in the CA1 and subiculum regions of the hippocampus).

The sample of 117 cases was subdivided into two classes: the first class included patients with a maximum number of NP in the hippocampus >1.5 (57 cases); the second class included patients with a maximum number of NP in the hippocampus <1.5 (60 cases).

The calculations clearly highlight four levels of results: the LDA in the random protocol obtained a mean accuracy equal to 65.10% (Arithmetical Mean); the FF-Sn ANNs obtained the best accuracy in the random phase with 76.70%; with the T&T and I.S. optimization protocol, the FF-Bp obtained 78.47% accuracy; and the SelfSABp Organism (Self-Recurrent Static Back Propagation) obtained the best result with an accuracy of 82.21%.

The variables selected by the I.S. system were 11 out of 16 (see Table 4, column E).

In this experiment, the chosen variables appear distributed over all the macro categories, with a prevalence for the variables belonging to the cognitive and linguistic functions category—all present. The R^2 values with respect to the target are the lowest of all the testing.

Classification of the Degree of Brain Atrophy in Neocortex

The degree of atrophy in the neocortex is a general measure of the degree of degeneration (atrophy) of the brain.

Gross examination of the participants' brains was performed by one neuropathologist who was blinded to the participants' health status and cognitive test scores. Following formalin fixation,

the intact brain was examined by the neuropathologist who rated the degree of atrophy of the frontal, temporal, and parietal lobes (the occipital lobes were rarely atrophic). The degree of atrophy was rated from 0 to 3, based on the degree of widening of the sulci and narrowing of the gyri in the three lobes. Severe atrophy (score of 3/3) was characterized by a high degree of widening of the sulci and narrowing of the gyri in two or more lobes; moderate and mild atrophy was characterized by varying degrees of the above in one or two lobes (1–2/3); and no atrophy was characterized by no noticeable widening of the sulci or narrowing of the gyri in any lobe (0/3).

The four stages (from 0 to 3) with which the brain's degree of atrophy was measured were grouped into two classes:

Absent/slight atrophy, 87 patients, equal to 74.36% of the sample.

Moderate/severe atrophy, 30 patients, equal to 25.64% of the sample.

The two classes make up the calculation models' target. Each patient is classified as belonging to one of the two classes by the different ANNs based on the 16 input variables.

Experimental design is similar to the above related predictive experiments. In this experiment three levels of results are identified: statistical systems with random samples, mean accuracy: 77.02%; ANNs with random samples (FF-Bp), mean accuracy: 80.65%; Organisms with Bp law of learning and sample optimization systems, mean accuracy: 90.65%.

The I.S. system selected 9 out of 16 input variables (see Table 4, column F).

In this last experimentation four variables out of nine belong to daily living activities, something which had never happened in the preceding experimentations: the variables which measure the cognitive and linguistic functions are only three. It therefore comes to light that the degree of cerebral atrophy seems to depend in a greater measure from the activities tied to daily life with respect to the other neurologic pathologies.

Prediction Experimentation

Considering the good predictive results obtained by the ANNs for the classification, we decided to carry out the following four supplementary experimentations:

- Estimation of the real number of NFT per mm^2 in the neocortex. Sample optimized by T&T and I.S.: 78 cases in Training and 39 in Testing, subsequently crossed over.
- Estimation of the real number of NFT per mm^2 in the hippocampus. Sample optimized by T&T and I.S.: 74 cases in Training and 43 in Testing, subsequently crossed over.
- Estimation of the real number of NP per mm^2 in the neocortex. Sample optimized by T&T and I.S.: 69 cases in Training and 48 in Testing, subsequently crossed over.
- Estimation of the real number of NP per mm^2 in the hippocampus. Sample optimized by T&T and I.S.: 75 cases in Training and 42 in Testing, subsequently crossed over.

The results shown are the average of the eight calculations with cross over.

Some statistical indexes were used for evaluating the good quality of the prediction (see Table 5). Indexes were calculated from the mean values of the NFTs truly present in the neocortex and in hippocampus, from the mean values of NPs truly present in the neocortex and in hippocampus, and from the values estimated by the calculation models.

Among the different cost function indexes available, the linear correlation (R) was used as main parameter to measure the predictive capacity of models. The results show that real mean

values of tangles and plaques present in the neocortex and in the hippocampus are highly correlated with the values estimated by the ANN.

In the four tests relating to prediction of Alzheimer pathology, the variables selected by the optimization system of the T&T and I.S. system were a subset of the 16 original ones. As shown in Table 6, the ApoE4 variable was selected by the T&T and I.S. systems in all the experiments, rendering it indispensable in the samples' optimization phase — a process which allows the ANNs and the Organisms to improve their predictive accuracy. This variable has a very low linear correlation coefficient with respect to all the targets. The relevance of this variable is confirmed by the results shown in Table 4, where it was selected five times out of six.

In the same way, three variables relative to cognitive and linguistic functions (MMSE, WRCL, and CNPR) were selected in all experiments. Among the variables used, these ones showed the highest degree of linear correlation.

Discussion

Our study shows that during life, ANNs can accurately predict the presence of AD pathology at death. The adaptive systems had superior accuracy when compared to models of traditional linear statistics.

The percentage of difference between the Organisms trained on the optimized samples and the discriminating analysis (LDA) ranged from 12.31% for NFT in the hippocampus, to 20.15% for NP in the neocortex. These findings are similar to the percentage of difference (17%) between LDA and Organisms in the forecast of Braak stages, subdivided into two classes (0–2 vs 3–6), and a difference of 13% between LDA and Organisms in the classification of the degrees of atrophy subdivided into two classes.

Further, accuracy by the Organisms was 20% better than LR in predicting the average number of tangles.

This study demonstrates that adaptive approaches based on ANNs are more suited to handle the nonlinear relationship of demographic, functional, cognitive, and genetic variables to neuropathologic lesions of AD. This approach is superior to traditional linear statistics and better suited to analyze the multifactorial elements, which may lead to the onset of AD.

Both NFTs and senile plaques are primary lesions of AD. Senile plaques include both diffuse plaques and NPs; usually the majority of senile plaques are the diffuse-type. However, there is a general agreement that NPs have a stronger correlation with cognitive impairment and dementia, than do diffuse plaques. For example, the CERAD neuropathologic criteria (Morris et al., 1989) for AD uses NPs in the neocortex to determine the likelihood that the brain is an “Alzheimer's brain.”

Our study shows that ANNs performed better in predicting the number of NFT in the neocortex and hippocampus than in predicting the number of NP in the same areas, either during the classification or in the value estimation testing process.

Cognitive and functional variables were better predictors for neocortical tangles (cut off: 1 tangle/mm²) than the Braak stages classification (0–2 vs 3–6). It is noteworthy that the cut-off determined through the use of adaptive systems corresponds to the threshold routinely employed by the pathologists of the Nun Study to determine the anatomically relevant degree of tangle burden.

The clinical variables selected predict with accuracy the number of NFT, but not NP, in each brain region. These results are consistent with recent findings (Arriagada et al., 1992a; Riley et al., 2002; Giannakopoulos et al., 2003; Guillozet et al., 2003) by other authors and our group suggesting that the presence of NFT in aging may represent one of its earliest pathological substrates and play a significant role in the initial stages of memory impairment. NFT formation in the neocortical area may represent a crucial step in the degenerative process because it may precede the emergence of the dementia. These results are further corroborated by a positive predictive outcome for severity of cortical atrophy.

In agreement with previous findings, our data suggest that additional factors could mediate the relationship between neuropathology and cognitive function. These factors may explain the variability found in cognitive state. The stronger association of predictive variables to NFT, rather than NP, implies an important role for NFT in the genesis of AD. Factors determining the formation of NFT in the neocortex may be more relevant to the clinical onset of AD than earlier pathological changes in the entorhinal cortex and hippocampus.

Our results indicate that a nonlinear analysis of complex data is a valid approach in shedding light on the role of NP and NFT in the development of a degenerative process leading to AD.

In all the experiments, the variables selected by the optimization system of the T&T and I.S. samples were a subset of the original ones. This indicates that there are some variables more suitable to predict a target with respect to another, while others provide information on different targets. The choice of variables with respect to the target is completely independent from their degree of linear correlation. Correlations are low for each variable. On the contrary, through the use of the T&T and I.S. systems, we know which variable is important for the determination of output, though it has a low or nonexistent degree of linear correlation.

The ApoE4 genotype is a good example of this concept. This variable showed a very low linear correlation with all targets. Notwithstanding, ApoE4 was selected by the I.S. system in all prediction experiments (Table 6), and optimized the predictive accuracy of ANNs.

Besides ApoE4 status, cognitive and linguistic functions (MMSE, WRCL, and CNPR) were consistently selected in all experiments by artificial organisms. Among the variables used, these showed the highest degree of linear correlation.

This study shows that the neural networks can detect complex interactions between variables and extract new medical information from raw data. They take into account many factors at the same time by combining and recombining the factors nonlinearly. Neural network outputs could be thought of as “composite variables” formed by nonlinear combinations of the originally input independent variables.

Conventional statistics reveal only a few parameters that are significant for the entire population, whereas ANNs handle all information which might not have significance for the entire population, but are highly predictive within subgroups or even for each individual patient. This is a qualitatively different approach compared with previous methods.

From a clinical point of view, the possibility to estimate the cerebral lesions load (both senile plaques and tangles) from simple descriptive clinical and demographic variables in individual patients make the results of this study extremely attractive. In fact when effective treatments targeted against neuropathological lesions undergo extensive clinical trials, the neural networks approach could provide medical decision support tools and a surrogate measurement of the expected reduction of Alzheimer pathology in individual patients.

In summary, the application of adaptive system analysis to clinical variables can offer a predictive test in complex clinical entities like AD, and at the same time provide us with significant pathophysiological clues in disease onset and progression.

Key elements for adaptive system analysis are:

- A complete baseline database of both number and quality of variables.
- A protocol for training the systems by placing them in the hardest environment possible to render the final predictive result more reliable.
- Optimization systems, which are able to extract samples and variables which are able to provide the greatest informative content from the data sets.
- Calculation models, which are able to explore the relationships between variables by using algorithms particularly suited to analyze complex phenomena.

The presence of all these elements allows the analysis of multifactorial pathologies such as AD. This is a radical divergence from the customary models of statistical linear calculation, which may be inadequate to treat such complex multifactorial problems, because of the poor nonlinear component.

At present we do not have methods to measure the presence of AD pathology in vivo. Imaging techniques, based on NFT and NP are currently under development but they have to face methodological issues, such as scanner spatial resolution in the relatively small brain regions, and development of a valid tangle-specific marker.

In this study, we have demonstrated that it is possible to train neural networks to predict the amount of tangles and amyloid plaques from the clinical descriptive data and neuropsychological tests. This approach can represent a noninvasive and inexpensive method for the assessment of AD specific brain lesion load in the presymptomatic phase of the disease and the monitoring of therapeutic interventions on NFT or NP.

Availability of the analysis code and data sources. The authors have made available the data and code requested. In particular:

- Analysis code, files of variables, and weight matrixes are available at Semeion Research Centre of Sciences of Communication, Rome, Italy.
- Data sources are available at Sanders-Brown Center on Aging, Department of Preventive Medicine, College of Medicine, University of Kentucky, Lexington, KY; and at the Medical College of Wisconsin 9200 West Wisconsin Avenue, Milwaukee, WI.

Acknowledgments

The authors thank Ms Gaia Gabbai for her support in typing and formatting the manuscript.

References

- Arriagada P, Growdon J, Hedley-Whyte E, Hyman B. Neurofibrillary tangles but not senile plaques parallel duration and severity of Alzheimer's disease. *Neurology* 1992a;42:631–639. [PubMed: 1549228]
- Arriagada P, Marzloff K, Hyman B. Distribution of Alzheimer-type pathologic changes in nondemented elderly individuals matches the pattern in Alzheimer's disease. *Neurology* 1992b;42:1681–1688. [PubMed: 1307688]
- Braak H, Braak E. Neuropathological staging of Alzheimer related changes. *Acta Neuropathol* 1991;82:239–259. [PubMed: 1759558]

- Bridle, JS. Probabilistic Interpretation of Feedforward Classification Network Outputs, with Relationships to Statistical Pattern Recognition. In: Fogelman-Soulié, F.; Héroult, J., editors. *Neuro-Computing: Algorithms, Architectures*. Springer-Verlag; New York: 1989.
- Buscema M. Artificial neural networks. (Special Issue Vol. I, Theory) *Subst. Use Misuse* 1998a;33(1)
- Buscema M. Artificial neural networks. (Special Issue Vol. II, Models) *Subst. Use Misuse* 1998b;33(1)
- Buscema, M. Reti neurali artificiali e sistemi sociali complessi. In: e modelli, Teoria; Angeli, Franco, editors. 1. Milano: 1999.
- Buscema, M. Supervised ver. 6.45. 12. Semeion Software; Rome: 1999–2003.
- Buscema M, Sacco PL. Feedforward networks in financial predictions: the future that modifies the present. *Expet. Syst* 2000;17(3):149–170.
- BuscemaMBredaMTerziS2000Sine NetRome Semeion Technical Paper, no. 21
- BuscemaMT&T: a new pre-processing tool for non linear dataset. 2001aRome Semeion Technical Paper, no. 25
- Buscema, M. I.S. (Input Selection) ver. 1.0. 17. Semeion Software; Rome: 2001b.
- Buscema, M. T&T (Training & Testing) ver. 1.0. 16. Semeion Software; Rome: 2001–2002.
- Buscema M. A brief overview and introduction to artificial neural networks, *Subst. Use Misuse* (Special Issue on the Middle Eastern Summer Institute on Drug Use, Proceedings. 2002;1997–1999;37(8–10):1093–1148.
- Buscema M. Genetic doping algorithm (GenD). theory and application. *Expet Syst* 2004;21(2):63–79.
- Dietterich T. Approximate statistical test for comparing supervised classification learning algorithms. *Neural Comput* 1998;10:1895–1923. [PubMed: 9744903]
- Giannakopoulos P, Herrmann FR, Bussiere T, et al. Tangle and neuron numbers, but not amyloid load, predict cognitive status in Alzheimer's disease. *Neurol* 2003;60:1495–1500.
- Giannakopoulos P, Hof P, Giannakopoulos AS, Herrmann F, Michel JP, Bouras C. Regional distribution of neurofibrillary tangles and senile plaques in the cerebral cortex of very old patients. *Arch. Neurol* 1995;52:1150–1160. [PubMed: 7492288]
- Guillozet AL, Weintraub S, Mash DC, Mesulam MM. Neurofibrillary tangles, amyloid, and memory in aging and mild cognitive impairment. *Arch. Neurol* 2003;60:729–736. [PubMed: 12756137]
- Mecocci P, Grossi E, Buscema M, et al. Use of artificial networks in clinic trias: a pilot study to predict responsiveness to Donepezil in Alzheimer's disease. *J. Am. Geriatr. Soc* 2002;50(11):1857–1860. [PubMed: 12410907]
- Morris JC, Heyman A, Mohs RC, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* 1989;39:1159–1165. [PubMed: 2771064]
- Petersen RC. Mild cognitive impairment: transition between aging and Alzheimer's disease. *Neurologia* 2000;15(3):93–101. [PubMed: 10846869]
- Price J, Davis P, Morris J, White D. The distribution of tangles, plaques, and related immunohistochemical markers in healthy aging and Alzheimer's disease. *Neurobiol. Aging* 1991;12:295–312. [PubMed: 1961359]
- Price J, Morris J. Tangles and plaques in nondemented aging and “preclinical” Alzheimer's disease. *Ann. Neurol* 1999;45:358–368. [PubMed: 10072051]
- Riley KP, Snowdon DA, Markesbery WR. Alzheimer's neurofibrillary pathology and the spectrum of cognitive function: findings from the Nun Study. *Ann. Neurol* 2002;51:567–577. [PubMed: 12112102]
- Snowdon DA, Kemper SJ, Morrimer JA, et al. Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: findings from the Nun Study. *JAMA* 1996;275:528–532. [PubMed: 8606473]
- Vomveg TW, Buscema M, Kauczor HU, et al. Improved artificial neural networks prediction of malignancy of lesions in contrast-enhanced MR-mammography. *Med. Phys.* Vol 2003;30(9):2350–2359.

Suggested Readings

- Buscema M. Self-reflexive networks. theory, topology, applications. *Qual. Quant. Int. J. Methodol* 1995;29(4):339–403.
- Davis D, Schmitt F, Wekstein D, Markesbery W. Alzheimer neuropathologic alterations in aged cognitively normal subjects. *J. Neuropathol. Exp. Neurol* 1999;58:376–388. [PubMed: 10218633]
- Petersen RC, Parisi JE, Johnson KA, et al. Neuropathological findings in patients with mild cognitive impairment. *Neurology* 1997;48:A102.
- Rumelhart, DE.; McClelland, JL. *Parallel Distributed Processing*. The MIT Press; Cambridge, MA: 1986.

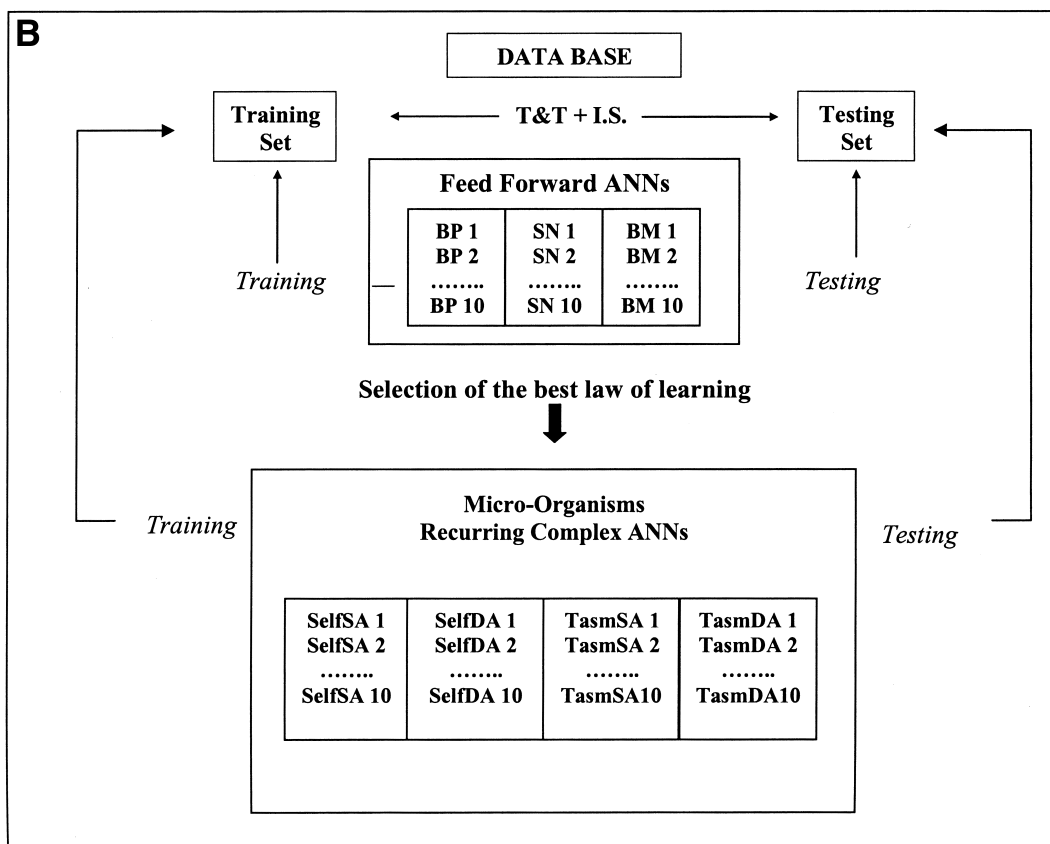
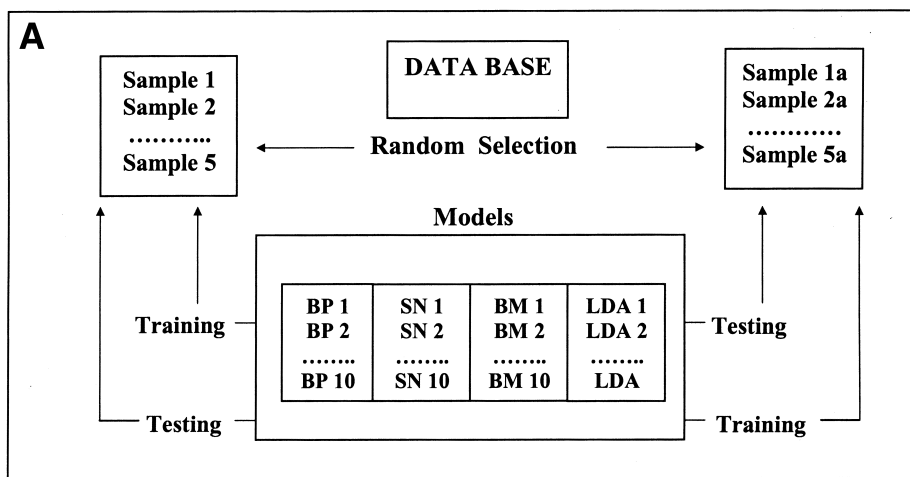


Fig. 1. Schematic representation of the research protocol. (A) Supervised ANNs – Research Protocol – Random (B) Supervised ANNs – Research Protocol – T&T + I.S.

Table 1

Database composition

Demographic variables

1. Age in years at last cognitive exam (Age)
2. Age in years at death (Age death)
3. Years of attained education (Educ-Yrs)

Global physical functions

4. Demonstrates ability to walk 50 ft at last exam (Walk)
5. Completes dressing tasks without help at last exam (Dress)
6. Demonstrates ability to stand up from a seated position at last exam (Stand)
7. Indicates ability to attend to toileting needs without assistance at last exam (Toilet)
8. Completes eating and drinking tasks without help at last exam (Eat-Drink)
9. Number of Daily Living Activities intact at last exam (ADL Tot)

Cognitive and linguistic functions

10. Delayed Word Recall score at last exam (WRCL)
11. Constructional Praxis score at last exam (CNPR)
12. Boston Naming score at last exam (BOST)
13. Verbal Fluency score at last exam (VRBF)
14. Mini-Mental State Exam score at last exam (MMSE)
15. Presence of a mild cognitive impairment with memory impaired during any of the first four exams (MCI)

Genetic variable

16. Presence or absence of one or more Apolipoprotein E4 alleles (ApoE4)
-

Table 2
 Values of the Linear Correlation (R^2) That Each of the 16 Variables Has With Its Respective Targets

Input Variables	Target of Each Experimentation					
	Braak	Mean Neurofibrillary Tangle Count in Neocortex	Mean Neurofibrillary Tangle Count in Hippocampus	Maximum Neuritic Plaque Count in the Neocortex	Maximum Neuritic Plaque Count in the Hippocampus	Degree of Atrophy in the Neocortex
Age	0.083	0.009	0.163	0.030	0.046	0.061
Age death	0.082	0.009	0.163	0.028	0.045	0.053
Educ-Yrs	0.004	0.006	0.006	0.008	0.014	0.040
Walk	0.010	0.058	0.023	0.013	0.001	0.165
Dress	0.050	0.074	0.058	0.042	0.019	0.119
Stand	0.027	0.075	0.043	0.027	0.005	0.167
Toilet	0.076	0.127	0.093	0.056	0.020	0.131
Eat-drink	0.072	0.137	0.079	0.055	0.048	0.186
ADL Tot	0.053	0.113	0.070	0.045	0.018	0.188
WRCL	0.382	0.253	0.283	0.189	0.114	0.123
CNPR	0.212	0.281	0.131	0.170	0.037	0.341
BOST	0.274	0.321	0.196	0.224	0.080	0.384
VRBF	0.246	0.287	0.234	0.155	0.049	0.204
MMSE	0.308	0.380	0.204	0.256	0.103	0.329
MCI	0.013	0.004	0.000	0.008	0.000	0.051
ApoE4	0.068	0.042	0.029	0.064	0.056	0.037

Table 3
 Prediction by Artificial Neural Networks and Linear Discriminant Analysis of the Neuropathological Outcomes by Using Cognitive and Functional Variables in 117 Subjects

Experiments	Models	Num	Target Classification		A. Mean (%)	W. Mean (%)
			Stages 0–2 (%)	Stages 3–6 (%)		
Braak stages	LDA–Random	10	74.00	70.72	72.36	71.91
	ANN FFSn–Random	30	83.58	76.14	79.86	79.72
	Bp Organisms Optimized	30	94.38	85.47	89.92	87.25
NFT in neocortex			Tangles < 1.0 (%)	Tangles > 1.0 (%)		
	LDA–Random	10	74.12	73.48	73.80	73.85
	ANN FFSn–Random	20	86.51	81.81	84.16	84.57
	Best ANN FFBp Optimized	10	89.18	86.49	87.83	87.08
	Organisms (SelfDABp)	40	93.36	87.37	90.36	89.95
NFT in hippocampus			Tangles > 10	Tangles < 10		
	LDA–Random	10	76.83%	74.21%	75.52	75.62
	ANN FFSn–Random	20	84.12	83.77	83.95	83.92
	Best ANN–FFBm Optimized	10	84.65	88.70	86.68	86.33
	Organisms (TasmSABm)	40	88.00	87.68	87.83	87.87
NP in neocortex			Plaques > 1.5 (%)	Plaques < 1.5 (%)		
	LDA–Random	10	64.03	75.50	69.77	66.37
	ANN FFSn–Random	20	87.11	80.25	83.68	85.16
	Best ANN–FFBp Optimized	10	89.74	88.13	88.93	88.11
	Organisms (SelfDABp)	40	88.59	91.25	89.92	88.62
NP in hippocampus			Plaques > 1.5 (%)	Plaques < 1.5 (%)		
	LDA–Random	10	61.68	68.53	65.10	65.36
	ANNs–FFSn Random	20	79.24	74.16	76.70	76.26
	Best ANN–FFBp Optimized	10	77.01	79.93	78.47	78.14
	Organisms (SelfSABp)	40	79.40	85.02	82.21	81.94
Atrophy			No-mild (%)	Moderate–Severe (%)		
	LDA–Random	10	88.00	66.04	77.02	82.54
	ANNs–Random	20	87.58	73.73	80.65	84.07
	Organisms	30	93.21	88.10	90.65	91.38

Num: number of elaborations; **A. Mean:** arithmetic mean; **W. Mean:** weighted mean.

Table 4

Classification Experiments

Input Variables	A						Total Presence of Each Variables
	Braak Stage	B	C	D	E	F	
Age	1	0	0	1	0	1	3
Age death	0	0	1	0	1	0	2
Educ-Yrs	0	0	0	1	0	0	1
Walk	0	1	1	1	0	0	3
Dress	1	0	0	1	1	1	4
Stand	1	1	0	1	1	0	4
Toilet	0	1	0	1	0	1	3
Eat-Drink	1	1	1	0	0	1	4
ADL Tot	0	0	0	1	1	1	3
WRCL	1	1	1	1	1	1	6
CNPR	0	0	0	1	1	0	2
BOST	0	1	0	0	1	1	3
VRBF	1	1	1	1	1	0	5
MMSE	1	1	0	1	1	0	4
MCI	0	0	0	1	1	1	3
ApoE4	1	0	1	1	1	1	5
Total number	8	8	6	13	11	9	

Variables selected by sample optimization systems in different experiments according to each target (0 = not selected; 1 = selected).

Table 5
 Prediction of Alzheimer Pathology: Results of the Four Experimentations

	Results	RMSE	Real Error	Absolute Error	Squared Corr.	Linear Corr.
NFT in neocortex	LR	0.171610	0.003387	0.161768	0.387576	0.621777
	FF-Sn	0.073882	-0.008056	0.062318	0.686611	0.824221
NFT in hippocampus	LR	0.398613	0.003193	0.433491	0.256982	0.504362
	FF-Bp	0.076190	0.006118	0.065249	0.694465	0.826696
NP in neocortex	LR	0.142933	-0.001360	0.162851	0.246499	0.492393
	FF-Bp	0.128231	0.006665	0.146139	0.354981	0.594993
NP in hippocampus	LR	0.161564	0.004348	0.173552	0.076107	0.266053
	FF-Bp	0.134843	-0.011556	0.134763	0.384627	0.608283

The linear regression is compared with the best ANNs.

Table 6

Prediction of Alzheimer pathology

Input Variables	NP in Neocortex	NP in Hippocampus	NFT in Neocortex	NFT in Hippocampus	Total Presence of Each Variable
Age	1	1	0	0	2
Age death	1	0	0	0	1
Education (yr)	1	1	0	0	2
Walk	1	1	0	0	2
Dress	0	0	0	1	1
Stand	0	1	1	1	3
Toilet	0	0	1	0	1
Eat-Drink	1	0	1	1	3
ADL Tot	1	0	0	1	2
WRCL	1	1	1	1	4
CNPR	1	1	1	1	4
BOST	1	1	1	0	3
VRBF	0	0	1	1	2
MMSE	1	1	1	1	4
MCI	1	1	1	0	3
ApoE4	1	1	1	1	4
Total number	12	10	10	9	

Variables selected by sample optimization systems in different experiments according to each target (0 = not selected; 1 = selected).