

# An Ensemble Approach for the Diagnosis of Cognitive Decline with Missing Data

Patricio García Báez<sup>1</sup>, Carlos Fernández Viadero<sup>2</sup>, José Regidor García<sup>3</sup>,  
and Carmen Paz Suárez Araujo<sup>3</sup>

<sup>1</sup> Department of Statistics, Operating Research and Computation, University of La Laguna,  
38271 La Laguna, Canary Islands, Spain

pgarcia@ull.es

<sup>2</sup> Unidad de Atención a la Dependencia de Santander, Gobierno de Cantabria, 39012 Santander,  
Cantabria, Spain

<sup>3</sup> Institute of Cybernetic Sciences and Technology, University of Las Palmas de Gran Canaria,  
35017 Las Palmas de Gran Canaria, Canary Islands, Spain

cpsuarez@dis.ulpgc.es

**Abstract.** This work applies new techniques of automatic learning to diagnose neuro decline processes usually related to aging. Early detection of cognitive decline (CD) is an advisable practice under multiple perspectives. A study of neuropsychological tests from 267 consultations on 30 patients by the Alzheimer's Patient Association of Gran Canaria is carried out. We designed neural computational CD diagnosis systems, using a multi-net and ensemble structure that is applied to the treatment of missing data present in consultations. The results show significant improvements over simple classifiers. These systems would allow applying policies of early detection of dementias in primary care centers where specialized professionals are not present.

**Keywords:** Cognitive Decline; Dementia; Diagnosis; Alzheimer's Disease; Cognitive Tests; Missing Data;  $k$ -fold Cross-validation; Artificial Neural Network; Modular Neural Network; Counterpropagation; Multi-net; Ensemble.

## 1 Introduction

The progressive aging of the world population has brought with it an increase in the incidence of diseases that are associated with this advanced age. Neurodegenerative processes, especially those including dementias, are two areas that are noteworthy. One outstanding feature is its high prevalence, 8-10% in individuals older than 65, and more than 25% of the population who are very advanced in years [1]. This prevalence has some very important repercussions in the patient's life and in other spheres. Its impact on the quality of life and individual autonomy is apparent, but we must not ignore the changes seen in the familiar, social and healthcare settings as well [2][3].

These consequences in these settings are a clear signal for research and applications to be carried out in possible precocious diagnosis. Of course the diagnosis has to be as accurate as possible while trying to resist the devastating effects of this pathology through the use of a therapeutic plan and suitable approach [4]. This is why the

early detection of CD is an advisable practice at every level of attention. However, the high uncertainty diagnosis [5], and the degree of the underdiagnosis which is so crucial and can reach 95% of the cases in some settings [6] highlights the need to develop detection programs and new instruments of diagnosis.

Three types of CD evaluation are possible. The first uses an open form, another is semi structured, while the third evaluation is carried out using the application of a series of cognitive evaluation scales. It must be stressed, however, that a test or a cognitive-based scale does not necessarily confirm deterioration or a normality diagnosis, but only indicates the possible existence of a cognitive deficit. Several neuropsychological tests [7] have been developed and used to evaluate different cognitive sections of a patient throughout the years. However, it is not a simple task to define a clear relationship between test results and specific symptoms or different levels of CD. Wide validation and testing are, of course, needed to determine levels of sensitivity, specificity and predictive value. Other important problems in the use of these scales are the absence of universal cut points, trans-cultural difficulties and the level of precision that seems to be similar with the use of short or long scales [8][9].

Automated decision making applications using consultation with *several experts* are now being rediscovered by the computational intelligence community. This application has emerged as a popular and heavily researched area, known as ensemble systems, and has produced favorable results when compared to results from single expert systems for a broad range of applications and under a variety of scenarios [10].

We present a proposal for CD diagnosis based on modular neural computation systems by means of a multi-net and an ensemble system that is directed towards the treatment of frequently missing data in physician-clinical environment.

## 2 Data Environment

The dataset includes results from 267 clinical consultations on 30 patients during 2005 at the Alzheimer's Patient Association of Gran Canaria [11]. The data structure includes a patient identifier, results from 5 neuropsychological tests, and a diagnosis of CD as well as differential dementia. An important advantage in this data set is in its homogeneity, each patient has scores from monthly tests, except the Mini Mental test, which is carried out twice a year. Nevertheless, even though the majority of the patients were tested 12 times, there are some patients with missing data in their consultations. In effect, we are working with a dataset where a missing data feature is present, since not all of the patients are subject to the complete set of tests.

A collection of 5 different data tests are used [11]: Mini Mental Status Examination (MMSE), Functional Assessment Staging scale (FAST), Katz's index (Katz), Barthel's index (Bar) and the Lawton-Brody's index (L-B).

The missing data from the MMSE test were introduced using interpolation from both annual tests. Almost all of the patients were subject to this test, and clinical experts agreed with this approximation. Still other missing values had to be accounted for, in this case 71 from a total of 1335. See Table 1 for a distribution by test and number of patients.

**Table 1.** Statistics of missing data in data set

<i>Test type</i>	<i>Number of missing data</i>	<i>Number of patients</i>
<i>MMSE</i>	36 (13.5%)	4 (13.3%)
<i>FAST</i>	0 (0%)	0 (0%)
<i>Katz</i>	33 (12.4%)	12 (40%)
<i>Barthel</i>	1 (0.3%)	1 (3.3%)
<i>Lawton-Brody</i>	1 (0.3%)	1 (3.3%)
<i>Total</i>	71 (5.3%)	16 (53.3%)

Different fields that make up successfully obtained information were preprocessed. This step was carried out to facilitate posterior convergence (in the use of the data set). The neuropsychological tests were standardized based on the minimum and maximum values that could be attained in these tests. Fields that were not completed were labeled as having missing values and were submitted to special treatment later.

Diagnostic values are used at a control stage. The diagnosis of the CD can result in one of four possible values: Without CD, Slight CD, Moderate CD and Severe CD. An analysis of the complete set of consultations reveals that 15% were diagnosed with Slight CD, 34% with Moderate CD and 51% with Severe CD. Recall that the data source does not include patients without CD. Regarding the diagnosis of dementias, 73% of the consultations correspond to Alzheimer-type dementias, while the remaining 27% were included in other types of dementias.

Data set size is one of the added difficulties which must be addressed when designing an artificial intelligence system which will aid in the diagnosis of DC. The data set limits how suitable the training of the used neural models will be, in addition to accurate generalization error. A method must be found that allows an efficient approach to this aspect of the problem. A cross-validation method is used and it improves training and error considerations [12]. These methods are based on making several partitions on the total data employed, or *resampling*. Diverse training is carried out with these data. A first part is used as a training set while the second one functions as a validation set. The generalization average error of the different estimations carried out on the different validation sets will provide a trustworthy measurement of the evaluated model error. The cross-validation variant was the denominated *k*-fold, and consisted in partitioning the data set into *k* subgroups, performing *k* training exercises, and leaving a validation data subgroup in each one while using the remaining (*k* - 1) as training data. The conducted partition on our queries was based on the identification of the patient involved in the consultation, that is, the consultations performed on the same patients were grouped in the same subgroup. Consequently we created 30 different subgroups, or a 30-fold consultation. This approach allows more objective results to be obtained because a model that has not been trained with other source consultations from the same patient will be used when the error validation of the consultation is carried out. In other cases the consultation using to train would have a much greater correlation with the evaluation consultation.

### 3 Diagnosis Systems

Two modular neural computing systems employing a multi-net and an ensemble approach [10] were used to approach the problem of CD diagnosis. The systems are based on modules that implement Counterpropagation neural networks (CPNs) [13]. The first system assembles these modules in a competitive way using a Decision Tree based Switching Ensemble (DTBSE) [14]. The desired goal is that the system chooses the module that, a priori, offers a more effective diagnosis in that type of consultation, based on available tests to diagnose a patient in a determined consultation. The second system uses a Simple Majority Voting Ensemble (SMVE) combining method to assemble the modules [10]. This system counts the outputs from the best modules as votes, and then selects as system output the class with the maximum number.

#### 3.1 Counterpropagation Modules

Counterpropagation architecture [13] is a modular neural network of two independent-learning cascaded layers. One layer is based on supporting the classification under self-organizing learning. It includes one first stage consisting of a quantifier layer with competitive learning, in our case a Kohonen self-organizing map (SOM) [15]. The second stage implements delta rule learning, since it receives a single input value of 1 from the previous stage. This second stage is equivalent to outstar learning [16]. One of the most important advantages in this network is its tremendous speed. Possible trainings take place between 10 and 100 times faster than the conventional back-propagation networks, producing similar results. Increased speed is attributed to the simplification which occurs in the self-organizing stage. Simplification also allows the second stage to use a simple classifier that produces proven convergence in linear problems, better generalization skills and a reduction in the consumption of computing resources.

The problem of missing input data in the first stage is avoided by using a variant of the SOM architecture that is able to process missing data [17]. This variant prevents missing values from contributing when coming out or modifying the weights. Even so, this way of approaching missing values is insufficient by itself, the proportion of missing values is excessive for this network because of its low tolerance.

#### 3.2 Decision Tree Based Switching Multi-net System

Our multi-net system is an alternative to the missing data processing of CPN. It treats the missing data training in all the CPN modules with different input subgroups that can be given as a variant of the 5 dimensions corresponding to each of the tests used during consultations when diagnosing a patient. The design of a DTBSE [14], adapted to the obtained results, causes the CPN module to be chosen because, a priori it more effectively diagnoses the input consultation according to the corresponding test.

The design of the DTBSE was based on the accomplishment of training and evaluation of the generalization error of the different CPN modules with every possible input combination configuration. The dominations between classifiers were obtained from the analysis of the obtained results. In other words, we considered that a classifier dominates another one if using an input subgroup used by the second one

produces a validation error equal to or less than the second one. Some modules were quickly discarded according to the obtained classifier that dominates and the suitable rules were designed to select at every moment the module with smaller generalization error according to the presence of tests in the input consultation.

### 3.3 Simple Majority Voting Ensemble System

Finally our DTBSE is not using the ability of CPN to process missing data, because the system only sends data with all available inputs to each module. The domination exclusion method also eliminates some of the better modules. The ability of our CPN modules to process inputs with missing data is incorporated into the design of a neural ensemble, SMVE. The neural ensemble uses the modules with the lowest average of validation errors and creates a simple majority voting process based on their outputs. The class with the maximum number of votes is selected as the system output. The combination of the outputs of several classifiers does not guarantee a superior performance to the best ensemble module, but it clearly reduces the risk of making a particularly poor selection [10].

## 4 Results and Discussion

Fig. 1 summarizes average error for the calculated training and validation sets by means of the cross-validation method for all of the CPN modules with different input subgroups. The validation error for each module was calculated using only the data without missing values according to the inputs of the module. It can be deduced that MMSE, FAST and Katz, in this order, are the most important tests when carrying out the diagnosis. The best module using all validation data is the CPN with the MMSE and FAST tests inputs (CPN MMSE+FAST). Average error reached in the validation sets are 8.99% and 7.44% in the training ones. Table 2 includes all the statistics for the validation sets.

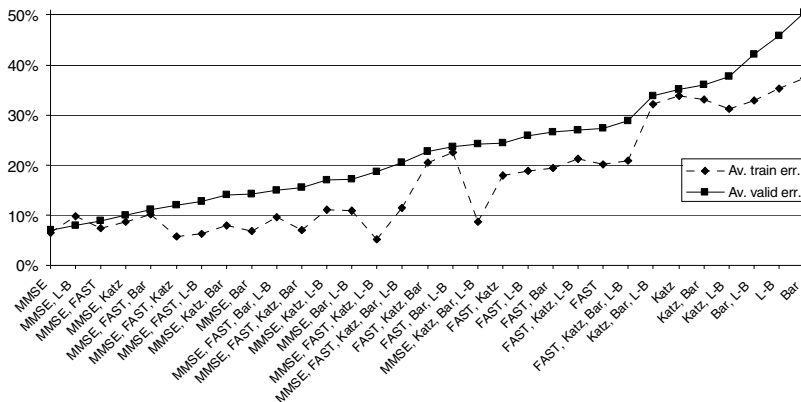
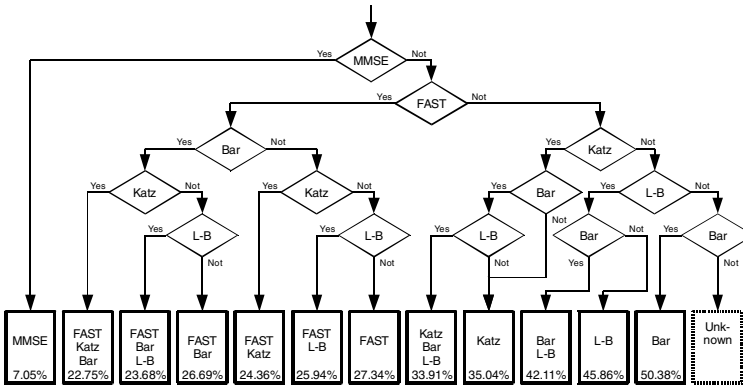


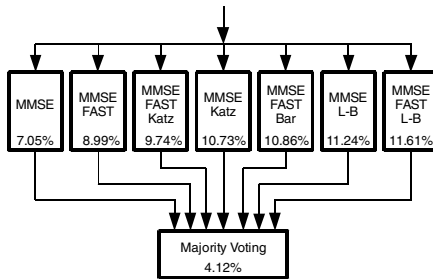
Fig. 1. Average training and validation errors for all the CPN modules with different inputs

The obtained results from Fig. 1 led to a decision to only use 12 of the most dominating modules. The DTBSE scheme used is described in Fig. 2. In it, depending on the not missing values present in the input consultation, the module with the best expected outcomes is selected.

The DTBSE obtained system presents an average error in the validation sets of 7.49%, Table 2 summarizes all the statistics for the validation sets. These results improve successfully, in a 1.5% the average validation error, to the best modules (CPN MMSE+FAST) applied on the total of the consultations, which reaffirms the utility of using a multi-net system such as the one described.



**Fig. 2.** DTBSE scheme and the modules used in it. Each module number is the validation error calculated using only the data without missing values according to the inputs of the module.



**Fig. 3.** SMVE scheme and the modules used in it. The number of each module is the validation error calculated using only the data with some available value to the inputs of the module.

The 7 best modules were used in the SMVE. The SMVE scheme used is described in Fig. 3. The class with the maximum number of votes is selected.

The SMVE obtained system with this approach presents an average error in the validation sets of 4.12% (see Table 2). These results successfully improve, by 4.87% the average validation error when compared to the best one from the modules (CPN MMSE+FAST) applied on the total of all consultations, and by 3.37% to the DTBSE. The drawback of this system is its computational requirements. Output computation of more than one module is required, unlike DTBSE, and combines the exit of theirs.

**Table 2.** Comparison between results for the validation sets of the three approaches

		<i>CPN MMSE+FAST</i>	<i>DTBSE</i>	<i>SMVE</i>
<i>Sensitivity</i>	<i>Slight CD</i>	59%	71%	76%
	<i>Moderate CD</i>	97%	93%	99%
	<i>Severe CD</i>	97%	99%	100%
<i>Specificity</i>	<i>Slight CD</i>	96%	85%	97%
	<i>Moderate CD</i>	81%	86%	90%
	<i>Severe CD</i>	100%	89%	100%
<i>Average error</i>		8.99%	7.49%	4.12%

## 5 Conclusions

The present work offers an important advance in the area of neuroinformatics and medical computer science based on a proposal to apply new techniques from automatic learning to diagnose cognitive decline processes usually related to aging.

The combination between the knowledge retrieved through cognitive tests and the ensemble techniques used to diagnose put our proposal in the field of Hybrid Intelligent Systems.

Required clinical tests may not be available at the time of a patient evaluation. This difficulty is overcome through the use of a variant of CPN with capacity for processing missing data. In addition, we also designed two multi-net systems that, according to missing input data, use the more suitable CPN modules during their treatment.

The use of this new instrument can help alleviate the degree of existing under-diagnosis, since it would allow policies of early detection of dementias to be applied in primary care centers where specialized professionals are not on staff. In addition the analysis conducted on the different modules that make up the system allows the best cognitive test for a correct diagnosis to be selected. Therefore the study of other possible tests could be extended to elaborate refined diagnosis protocols.

**Acknowledgments.** We would like to thank the Gobierno de Canarias for their support under Research Project PI2003/178A. Additional thanks are extended to the Alzheimer's Patient Association of Gran Canaria for their generous input of data.

## References

1. López-Pousa, S.: La demencia: concepto y epidemiología. In: Enfermedad de Alzheimer y otras demencias, pp. 33–42. Médica Panamericana, Madrid (2006)
2. Andersen, C.K., Wittrup-Jensen, K.U., Lolk, A., Andersen, K., Kragh-Sorensen, P.: Ability to perform activities of daily living is the main factor affecting quality of life in patients with dementia. *Health Qual Life Outcomes* 2, 52 (2004)
3. DeKosky, S.: Early intervention is key to successful management of Alzheimer disease. *Alzheimer Disease and Associated Disorders* 17, S99–S104 (2003)
4. Doody, R.S.: Refining treatment guidelines in Alzheimer's disease. *Geriatrics* 60, 14–20 (2005)

5. Neuropathology Group of the Medical Research Council Cognitive Function and Ageing Study: Pathological correlates of late-onset dementia in a multicentre, community-based population in England and Wales. *The Lancet* 357, 169–175 (2001)
6. Solomon, P., Murphy, C.: Should we screen for Alzheimer's disease? *Geriatrics* 60, 26–31 (2005)
7. Nestor, P.J., Scheltens, P., Hodges, J.R.: Advances in the early detection of Alzheimer's disease. *Nature Reviews Neuroscience* 5, S34–S41 (2004)
8. Borson, S., Scanlan, J.M., Watanabe, J., Tu, S.P., Lessing, M.: Simplifying detection of cognitive impairment: comparison of the Mini-Cog and Mini-Mental State Examination in a multiethnic sample. *Journal of the American Geriatrics Society* 53, 871–874 (2005)
9. Damian, M., Kreis, M., Krumm, B., Hentschel, F.: Optimized neuropsychological procedures at different stages of dementia diagnostics. *Journal of the Neurological Sciences* 229, 95–101 (2005)
10. Polikar, R.: Ensemble Based Systems in Decision Making. *IEEE Circuits and Systems Magazine* 6, 21–45 (2006)
11. García Báez, P., Suárez Araujo, C.P., Fernández Viadero, C., Regidor García, J.: Automatic Prognostic Determination and Evolution of Cognitive Decline Using Artificial Neural Networks. In: Yin, H., Tino, P., Corchado, E., Byrne, W., Yao, X. (eds.) *IDEAL 2007*. LNCS, vol. 4881, pp. 898–907. Springer, Heidelberg (2007)
12. Hjorth, J.S.U.: *Computer Intensive Statistical Methods Validation, Mod. Sel., and Bootstrap*. Chapman & Hall, London (1994)
13. Hecht-Nielsen, R.: Counterpropagation networks. *Applied Optics* 26, 4979–4984 (1987)
14. Sharkey, A.J.C.: Multi-Net Systems. In: *Combining Artificial Neural Nets*, pp. 1–30. Springer, Heidelberg (1999)
15. Kohonen, T.: *Self-Organization and Associative Memory*, 3rd edn. Springer Series in Information Sciences, Berlin (1989)
16. Grossberg, S.: Embedding fields: A theory of learning with physiological implications. *Journal of Mathematical Psychology* 6, 209–239 (1969)
17. Samad, T., Harp, S.A.: Self-organization with partial data. *Network* 3, 205–212 (1992)