Toward an intelligent computing system for the early diagnosis of Alzheimer's disease based on the modular hybrid growing neural gas

Ylermi Cabrera-León[1] (iD), Pablo Fernández-López[1] (iD), Patricio García Báez[2],
Konrad Kluwak[3] (iD), Juan Luis Navarro-Mesa[4], Carmen Paz Suárez-Araujo[1]
and for the Alzheimer's Disease Neuroimaging Initiative[*]

Abstract

Objective: The proportion of older people will soon include nearly a quarter of the world population. This leads to an increased prevalence of non-communicable diseases such as Alzheimer's disease (AD), a progressive neurodegenerative disorder and the most common dementia. mild cognitive impairment (MCI) can be considered its prodromal stage. The early diagnosis of AD is a huge issue. We face it by solving these classification tasks: MCI-AD and cognitively normal (CN)-MCI-AD.

Methods: An intelligent computing system has been developed and implemented to face both challenges. A non-neural pre-processing module was followed by a processing one based on a hybrid and ontogenetic neural architecture, the modular hybrid growing neural gas (MyGNG). The MyGNG is hierarchically organized, with a growing neural gas (GNG) for clustering followed by a perceptron for labeling. For each task, 495 and 819 patients from the Alzheimer's disease neuroimaging initiative (ADNI) database were used, respectively, each with 211 characteristics.

Results: Encouraging results have been obtained in the MCI-AD classification task, reaching values of area under the curve (AUC) of 0.96 and sensitivity of 0.91, whereas 0.86 and 0.9 in CN-MCI-AD. Furthermore, a comparative study with popular machine learning (ML) models was also performed for each of these tasks.

Conclusions: The MyGNG proved to be a better computational solution than the other ML methods analyzed. Also, it had a similar performance to other deep learning schemes with neuroimaging. Our findings suggest that our proposal may be an interesting computing solution for the early diagnosis of AD.

Keywords

Alzheimer's disease, artificial neural network, computer-aided diagnosis, mild cognitive impairment, machine learning

[1]Instituto Universitario de Cibernética, Empresa y Sociedad, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain
[2]Departamento de Ingeniería Informática y de Sistemas, Universidad de La Laguna , San Cristóbal de La Laguna, Spain
[3]Department of Control Systems and Mechatronics, Wrocław University of Science and Technology, Wrocław, Poland
[4]Instituto Universitario para el Desarrollo Tecnológico y la Innovación en Comunicaciones, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain
[*]Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.

usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Corresponding author:
Carmen Paz Suárez-Araujo, Instituto Universitario de Cibernética, Empresa y Sociedad, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain. Email:carmenpaz.suarez@ulpgc.es

Introduction

Rising older populations generates increased chronic and non-transferable illnesses, such as dementia and stroke, posing important socio-economical challenges.[1] Hence, dementia is considered a public healthcare issue, especially in countries with high life expectancy, as it is estimated that around 47 million people have dementia worldwide.[1]

Alzheimer's disease (AD), a progressive and irreversible neurodegenerative syndrome that produces dementia, accounts for approximately 70% of the global dementia cases[2–4]: an estimated 6.7 million have AD in the USA in 2023, rising to 13.8 million in 2060.[5] While aging is recognized as a very important risk factor of AD,[2] the exact etiology, development, and evolution of AD are not fully understood yet. Nevertheless, two pathological changes, amyloid plaques, and neurofibrillary tangles, have formed the most prominent lines of research trying to explain the etiological mechanisms of AD.[2] AD severely impairs daily activities and emotional states.[3]

Researchers have been looking for biomarkers that allow the, preferably early, diagnosis and prognosis of AD. The main difficulty of these tasks is how similar AD symptoms are to those of other diseases, especially other types of dementia. So far, it cannot be said that there is a specific biomarker for AD that is accepted by the scientific and medical communities and is of standardized use, but several promising candidates have been discovered. Among them, Blennow and Zetterberg pointed out the neurogranin, a synaptic protein apparently specific for AD that may allow predicting future cognitive impairment.[6] Similarly, González-Sánchez et al.[7] concluded that lactoferrin, an antimicrobial peptide commonly found in saliva, has high sensitivity and is specific for AD diagnosis, as it was not found in cognitively normal (CN) and frontotemporal dementia subjects. Unfortunately, none of these promising biomarkers are among the most popular, a position occupied by neuroimaging techniques.[8] Furthermore, currently, there is no effective treatment for AD. Nevertheless, early interventions can enhance life quality for patients and caregivers, often family members or non-clinicians.[2,4]

MCI, marked by slight cognitive deficits without dementia and without affecting daily activities, may presage AD but does not always progress to it. Roughly 5% to 10% of MCI cases transition to dementia annually, 10% to 15% to AD.[9,10]

Because of these considered aspects and the dire consequences of AD prevalence, research into early AD and differential diagnosis is an absolute requirement.[2,9] The lack of standardized diagnostic criteria turned the AD and MCI diagnosis into a complex problem, generating an important underdiagnosis.[3,11,12]

Our studies in this field are based on neural computation approaches. Our proposal applies a hybrid artificial neural network (ANN) to two classification tasks, a binary and a multiclass one: the differential diagnosis of AD and MCI, and the separation of CN, MCI, and AD subjects.

Many studies have employed deep learning (DL) and different ANN for facing this kind of binary classification, but typically using neuroimaging, such as magnetic resonance imaging (MRI) and positron emission tomography (PET).[14,17,13,15,16,18] Less common are both the multiclass models[19–21] or the prognosis or longitudinal approaches, where the progression of CN subjects to MCI or from MCI to AD is studied.[22,23] In recent years, there is a predominant use of deep neural network (DNN),[20,24,17,16,18] but also a great variety of ANN[26,28,27,25] or even other mathematical methods[29] have been proposed. While effective, they often rely on invasive or costly criteria, limiting their primary care usage.[24] Approaches based on non-neuroimaging techniques normally have similar performance but have been less frequent in the last decade.[8]

The main goal of this paper is twofold: to provide intelligent and effective computational solutions to aid in the classification of not only AD versus MCI subjects, but also CN versus MCI versus AD subjects. Our study presents a hybrid neural architecture called modular hybrid growing neural gas (MyGNG) for both classification tasks. This intelligent system facilitates early diagnosis and clinical decision-making across settings, particularly in primary care.[27] Finally, for each of these tasks, a comparative study of the performance of our ontogenetic neural architecture and several popular neural and non-neural machine learning (ML) models was included.

Dataset and method

Dataset

As two different classification problems were tackled in this work, two datasets were required, which will be described separately. They were built with data from the ADNI database.[1] Since 2003, ADNI, as a public–private partnership, has been providing a huge database where many medical tests from different patients have been collected over long periods of time. At the same time, ADNI has been studying comprehensively AD-related omics and imaging.[30] The main goals of ADNI are the early-as-possible diagnosis of AD, and to help improving its prevention, intervention, and treatment by finding new and reliable diagnostic techniques.[27]

In order to facilitate more adequate comparisons with some previous works,[27,25] the same dataset was used for the MCI-AD binary classification task. The first dataset comprised 345 MCI and 150 AD patients, that is, a total of 495 subjects, who started their participation in ADNI from the ADNI2 study. The "ground truth" label of these participants was given by clinicians following an exhaustive diagnosis criterion indicated by ADNI. Data from the baseline were utilized. The number of characteristics extracted for each subject was 211, which included demographic information, neuropsychological tests and their items, brain measurements obtained via MRI and PET, genetics, and other biomarkers.

The second dataset used in this work comprised 229 CN, 402 MCI, and 188 AD subjects, that is, a total of 819 patients. These participants belonged to the ADNI1 study and, similar to those in the previous dataset, they met an exhaustive diagnosis criterion given by ADNI. Data from the baseline were also used. An identical number of characteristics of the same modalities was extracted for each subject.

Method

Both classification tasks related to the early diagnosis of AD have been addressed in this work with neural computing methods, more specifically based on the GNG,[31] an ontogenic ANN. Ontogenetic neural architectures are ANN able not only to modify their connections during learning, as the rest of ANN, but also to automatically adjust their topology to the problem.[32,33,25] Due to these characteristics, they are quite suitable for clustering, vector quantization and data visualization.[33,25]

In this work, we have used a two-module hybrid neural architecture named MyGNG, which is a simpler and improved version from the one that was introduced in Sosa-Marrero et al.[27] The MyGNG has two main modules, Figure 1: the first one is built with an unsupervised, self-organizing, and ontogenetic module, which was based on Fritzke's GNG,[31] whereas the second module was based on a supervised neural architecture as is the perceptron.[34,35]

MyGNG is hierarchically organized the way it is on purpose: first the data clustering and later the data labeling.

The reason for doing the clustering, which other classifiers do not do, is to simplify the input space by projecting it to other reduced dimensions while preserving its topology in order to ease the classification done later by the labeling module. Apart from the expected increase in performance, it may also bring along decreased training times, as it also happens in other hybrid architectures such as the counter-propagation network.[19]

These modules learn sequentially. That is, the training of the MyGNG is done in cascade: the second module (named "Supervised" in Figure 1) uses for its training the labels of the data and the output of the first module (named "GNG") after this one has been trained.

In Figure 1, the structure of this improved MyGNG is depicted, with the input layer and its two sequential modules. The colors of the neurons indicate where two biologically related processes have happened (blue is the base one, and it is used for neurons that are adapting to the input data): neurogenesis and neural apoptosis. Green neurons represent new neurons, that is, they have been recently created (i.e. neurogenesis) where the GNG algorithm is considered more convenient (i.e. between the neuron with the greatest error and its neighbor with the greatest error). These green neurons require that the old connections be deleted (red lines) and new connections be created (green dashed lines). Conversely, red neurons are those that have been removed (i.e. neural apoptosis), which occur after all connections to them have been deleted.

The main difference between the improved MyGNG presented in this work and the original one in Sosa-Marrero
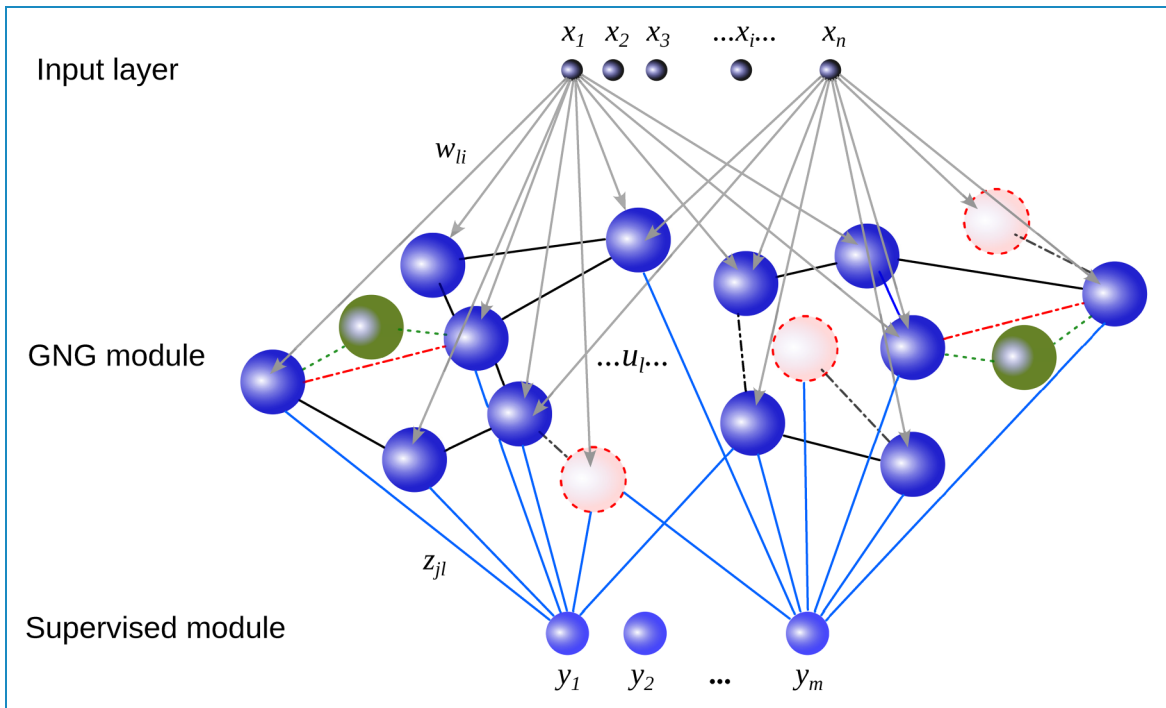
Figure 1. Structure of the MyGNG, where x_i is the ith component of the input vector; w_{li}, the weight of neuron li; and u_l, a neuron/unit of the GNG module. MyGNG: modular hybrid growing neural gas; GNG: growing neural gas.

et al.[27] is how the "Supervised module" is built, which will be explained later in (4). This module has now less complexity as it is based on a perceptron instead of the complex "Supervised module" of the original MyGNG, which made use of neural neighborhoods.[27] These neural neighborhoods are unnecessary and, hence, not used in the improved MyGNG introduced in the current work.

Regarding the first module of the MyGNG presented in this work, in Fritzke,[31,33] the GNG is described as a self-organization map based on a dynamic graph of connected neurons. Starting from a low number of interconnected neurons, this graph will adapt, shrink, and grow, hence producing topological learning that will allow clustering of the input space. This generation and continuous update of the graph is made by a competitive learning algorithm,[36] where the winner neuron[2] s_1 is the one whose weights (ω) are the most similar to the input vector ξ. Equation (1), where ε_b and ε_n are the learning rates for the winner and its neighbors, respectively,[31] indicates that the adjustment of the winner neuron and its direct topological neighbors defines the adaptation process.

$$\Delta\omega_{s_1} = \varepsilon_b(\xi - \omega_{s_1})$$
$$\Delta\omega_n = \varepsilon_n(\xi - \omega_n) \text{ for all direct neighbor } n \text{ of } s_1 \quad (1)$$

It should be noted that only the GNG equations that have been considered the most relevant are explained in this work. For the full list of equations, we recommend Fritzke's cited works.[31,33] These most relevant parts are related to the two processes that make GNG stand out from other ontogenic neural architectures: neurogenesis and neural apoptosis.

A local error variable is calculated for the winner neuron in each iteration (2). This error is related to the neurogenesis (i.e. the creation of a neuron) process because it allows identifying regions where the input signals are not sufficiently correctly represented. That is, a new neuron needs to be inserted between the unit q with the maximum error and its neighbor f in the graph that has the highest error, (3). An insertion occurs in every λ adaptation step. Error variables of these units q and f are reduced in proportion to the parameter β.

$$\Delta\text{error}(s_1) = \|\omega_{s_1} - \xi\|^2 \quad (2)$$

$$\omega_r = 0.5(\omega_q + \omega_f) \quad (3)$$

Altering connections modifies the network topology. A new connection is created on each adaptation step between the winner and the second winner neurons. Conversely, a connection is removed when the value of its age property is above the $a_{\max}$ parameter. Neural apoptosis (i.e. the deletion of a neuron) occurs when it becomes isolated after all the connections to that neuron are erased.

The responsible of the hybrid character of this MyGNG is the addition of a monolayer-perceptron-based output module (supervised learning) after the GNG-based one (unsupervised learning). The learning process of the perceptron is given by the "Perceptron rule" shown in (4), which indicates how the weights are updated.[35] In this equation, $x(k)$ is an input; $\omega(k)$, weights; ρ, the learning rate; and $\tilde{e}(k) = d(k) - y(k) = d(k) - \text{sgn}[\omega^T(k) \cdot x(k)]$, being sgn the sign function; $d(k)$, the desired output value of the perceptron for the input $x(k)$; and $y(k)$, the obtained output value of the perceptron for that input.

$$\omega(k+1) = \omega(k) + \rho \cdot \frac{\tilde{e}(k)}{2} \cdot x(k) \quad (4)$$

The following algorithm describes how the MyGNG works, where the parts regarding the GNG were loosely based on Fritzke's GNG algorithm[31]:

1. Create two neurons, a and b, with weights ω_a and ω_b, respectively.
2. Extract a sample ξ from the input space or, alternatively, generate an input signal according to the probability density function $P(\xi)$.
3. Find s_1 and s_2, the two neurons that are the nearest to the input sample.
4. All connections to s_1 have their *age* property incremented.
5. Increment the local error variable of s_1 (2). These errors are used to find where to insert a new neuron.
6. Move s_1 closer to ξ (1). Similarly, move all the direct neighbors of s_1 but by a lesser amount.
7. The age of the connection between s_1 and s_2 is reset to 0. Create it if it did not exist.
8. Neural apoptosis: after removing all connections whose *age* is greater than a_{max}, delete all neurons without connections.
9. Neurogenesis (3): every λ iterations, insert a new neuron between the neuron with the largest error and its direct neighbor with the largest error. The connection between the erroneous neurons is deleted, and two new connections are created: between each of them and the new one. The error of the erroneous neurons is diminished.
10. Decrease all error variables.
11. Go to Step 2 if the stopping criterion (e.g. epochs, performance metric, size of the network, etc.) of the GNG is not met yet. In our case, it is the epochs or number of times all the input samples are used for training the GNG.
12. Obtain an output from the GNG and the associated class label (i.e. the expected output of the perceptron).
13. Update the weights of the perceptron (4).
14. Go to Step 12 if the stopping criterion of the perceptron is still not met. In our case, it is the epochs used for training the perceptron.

Regarding the computational complexity of the MyGNG, O_{MyGNG}, as a modular ANN, it is the sum of the

computational complexities of each of its modules (5). The computational complexity of the first module is that of the GNG, O_{GNG}, which has been calculated by Mendes et al.[37] and is a function of k, the size of the input to the GNG (3 or 4 in the current work, i.e., the number of principal components (PCs)), and n, the number of neurons in the GNG. Their calculation assumes that: (a) the number of neighbors is kept low (lower than n) and (b), the quantity of neurons that are deleted is low (in order to consider that a graph with n neurons is created in λn iterations, where λ is the GNG parameter that indicates the frequency of neurons insertions). Both assumptions are met in our implementation. On the other hand, the computation complexity of the perceptron, $O_{\text{Perceptron}}$, is a function of the previous k and C, the number of neurons in the perceptron (2 or 3, that is the same as the number of classes in the classification tasks).

$$O_{\text{MyGNG}} = O_{\text{GNG}} + O_{\text{Perceptron}} = O(kn^2) + O(kC) \quad (5)$$

where $k = 4, n = 75$, and $C = 3$ for the worst-case scenario in this work.

MyGNG for the early diagnosis of AD

The intelligent hybrid system presented in this work is formed by two stages: a preprocessing one followed by a processing one. This system is considered hybrid because the first stage is non-neural whereas the second one is based on an ANN. The first stage comprised the next processes: imputation of missing values, feature ranking, data scaling, and data projection. The imputed, scaled, and projected data derived from the features that were ranked formed the input of the processing stage, which was based on the MyGNG.

Our system was mainly implemented with Python 3.10 and Tensorflow 2.10. Some preprocessing steps were implemented with "scikit-learn," a library for ML.[38]

In this work, two classification problems related to the early diagnosis of AD were studied, which will be analyzed separately in the "MCI versus AD" and "CN versus MCI versus AD" sections. For the global assessment of the intelligent computing system, the most common performance metrics in medicine, such as specificity, sensitivity (or recall), precision, and accuracy, were used.[28] Due to their adequacy for clinicians, both variants of the less popular metric clinical utility index (CUI) were also utilized[39]: CUI+ and CUI−. Receiver operating characteristic (ROC) curves were also used but only for comparisons with other ML algorithms. AUC, which is based on the ROC curve, was considered as the main performance metric due to not being affected by class unbalanced datasets.[40]

In each of these two classification tasks eight different scenarios were studied, which asked the following three questions: "is the addition of AGE, a demographic data

which is considered a risk factor in AD,[2] to the feature set beneficial to improve the classification performance?", "is it better to project the data with 3 or 4 components?", and "does scaling these datasets in the robust way provide an advantage over using the standard one?"

MCI versus AD

Before the input data were provided to our neural architecture and the other ML models which were later compared with, data required to be conveniently preprocessed. Imputation of missing values, feature ranking, data scaling, and data projection were four processes that needed to be carried out on our data before the processing stage. Thanks to them, the final datasets lacked missing values, the scales of all the features were in similar ranges, and their number decreased. This way, the original datasets were represented by a higher quality subset of features that will allow the models to require less training times and achieve better performance results.[41] As these preprocessing steps were not exactly identical for each classification problem, they will be described separately.

Missing values are generally considered a big handicap for most ML models. Due to the presence of missing values in several patients and attributes of our dataset, some imputation methods were analyzed: leaving the missing values untouched, discarding the participants with missing values in one of the used features, substituting the missing values with the median value for the class of that sample, idem but with the mean, and similar but with the mode. Imputing with the median value per class was deemed more robust and also produced the best results so it was finally selected for further usage.

Feature selection or ranking was performed due to the large number of attributes per subject, 211, in the MCI-AD dataset. As missing values were handled, this number remained identical for all participants. Two techniques were evaluated: Extreme Gradient Boosting (XGBOOST)[42] and fast correlation-based filter (FCBF).[43]

XGBOOST is a scalable tree ensemble method that, as a byproduct, also generates a ranking of the features. Unlike FCBF, redundant features are never discarded internally, which results in rankings including more than one feature providing similar information. Hence, feature ranking with XGBOOST was considered of lower quality, and the one done with FCBF was preferred.

In Yu and Liu,[43] the hybrid filter and wrap feature selection method named FCBF was introduced. Based on symmetric uncertainty, which finds not only the correlation between features and categories but also the redundancy between the features, FCBF only selects features that are highly correlated to categories and, at the same time, lowly correlated to other features. This way, calculation efficiency, and hence speed, is enhanced, thus improving the recognition rate.[27] In Yu and Liu,[43] FCBF demonstrated its good ability to identify redundant features in several

difficult datasets after being able to reduce their dimensionality more than other methods.

A vector of the most adequate features for the MCI-AD classification task according to FCBF (i.e. those with the highest FCBF scores) was obtained. Thanks to this, the number of features was vastly reduced from 211 to 6, Table 1 and Figure 2. These six features derived from three neuropsychological tests: four from the mini-mental state examination (MMSE),[44] one from the Alzheimer's disease assessment scale-cognitive subscale,[45] and the last one from the functional activities questionnaire (FAQ).[46]

Scaling of the data was tested by applying several methods: not scaling the data, standard scaling (by removing the mean and scaling the data according to the standard deviation; the most popular method), and robust scaling (similarly, but with the median and the interquartile range, respectively, more robust to outliers). Unlike in Sosa-Marrero et al.[27] and Cabrera-León et al.,[25] where using no scaling method was preferred, in our case both standard and robust scaling were studied and, as shown

Table 1. Characteristics of the subjects: A demographic feature, and the six attributes used by the model as input, sorted according to their FCBF score.

	AD	MCI
Number of subjects	150	345
AGE: mean	74.67	71.56
AGE: StD	8.18	7.38
MMSCORE: mean	23.07	27.98
MMSCORE: StD	2.08	1.74
MMDATE: mean	1.6	1.08
MMDATE: StD	0.49	0.26
MMBALLDL: mean	1.67	1.14
MMBALLDL: StD	0.47	0.35
ADAS_Q7: mean	2.4	0.43
ADAS_Q7: StD	1.75	0.85
MMYEAR: mean	1.27	1.0
MMYEAR: StD	0.44	0.05
FAQSHOP: mean	2.78	0.5
FAQSHOP: StD	1.82	1.14

FCBF: fast correlation-based filter; AD: Alzheimer's disease; MCI: mild cognitive impairment; StD: standard deviation.

later, they provided similar performance, albeit the former was more beneficial for the MyGNG.

Unlike in Sosa-Marrero et al.[27] and Cabrera-León et al.,[25] where principal component analysis was applied,[48,47] in this work neighborhood component analysis (NCA) was utilized for data projection, although it can be used for classification too.[49] NCA is a supervised method aimed at finding the best input data projection or linear transformation for a stochastic nearest neighbors rule to yield the best classification accuracy in the transformed space, without assuming that the data have a parametric structure in the low-dimensional representation.[49] Among the methods for the initialization of the linear transformation in the NCA that were analyzed, "identity" was considered the most adequate option because similar ranges of values of the components were obtained in all the eight scenarios that were studied.

The stratified K-folds cross-validation method, with five folds as it kept the same training-test subsets proportion used in Sosa-Marrero et al.[27] and Cabrera-León et al.,[25] was utilized for data partitioning. Its main advantage is that it keeps an identical proportion of samples for each class in all the folds. It was used by not only the MyGNG, but also the ML models studied in the "Comparative studies with ML models" section.

Several sets of values were given to the hyperparameters of the MyGNG in order to find the optimal combination. Two combinations were found of interest: one when 3 PCs were used and one when 4 PCs were used. For each of them, the same combination was found optimal when AGE was added to the feature set. Hence, for the 4 PCs scenario, with and without the AGE feature, the values of the hyperparameters for the GNG module were: training during eight epochs, 75 was the maximum number of neurons, five was the maximum age of any connection, 50 was the number of iterations before a new neuron is created, the learning rates $\varepsilon_b = 0.5$ and $\varepsilon_n = 0.015$, and the decremental parameters $\beta = 0.6$ and $d = 0.7$. Its values of the hyperparameters for the perceptron module were $\rho = 0.01$ and trained during 50 epochs. On the other hand, for the 3 PCs scenario, a similar combination was found optimal, which only differed in 3 hyperparameters of the GNG module: $\varepsilon_b = 0.65$, $\beta = 0.7$ and $d = 0.8$. In Table 2 the performance results of the MyGNG in each of the eight scenarios that were studied are shown.

In order to do a brief qualitative comparative study, several articles from other authors that dealt with the MCI-AD binary classification task and made use of data from ADNI were selected from the existing literature, Table 3. It showed that our MyGNG sometimes yielded better performance results than proposals from other authors, despite not only using non-expensive, non-invasive, non-ionizing, and easily applicable diagnostic criteria, such as neuropsychological scales, but also the MyGNG not being based on DNN, whose performance is
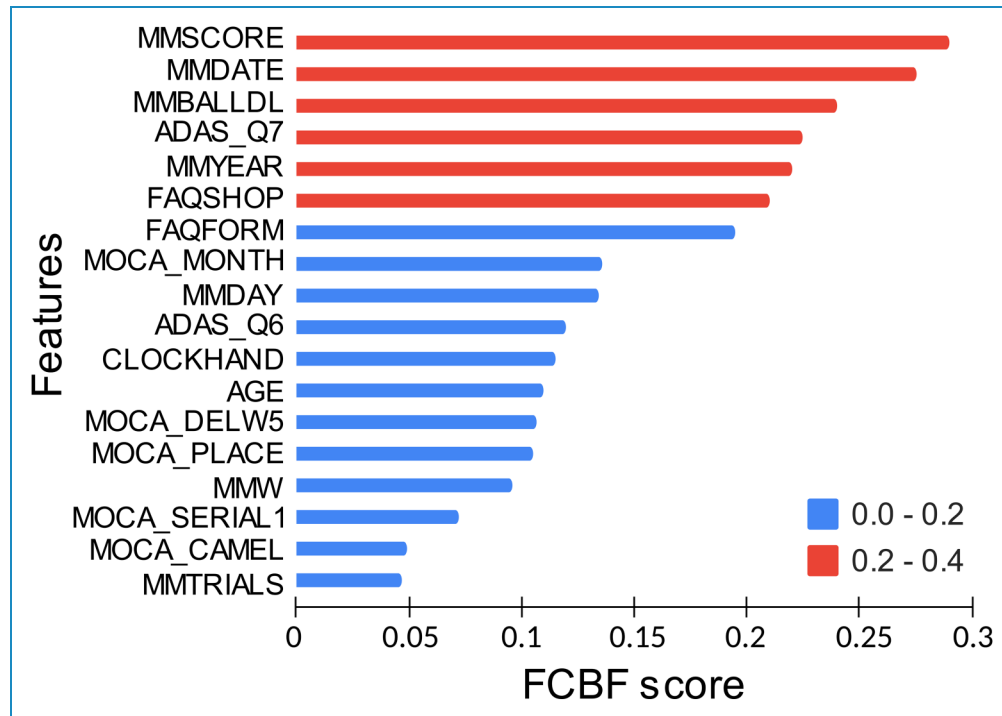
Figure 2. Ranking of features according to the FCBF method for MCI and AD subjects from ADNI2. FCBF: fast correlation-based filter; MCI: mild cognitive impairment; AD: Alzheimer's disease; ADNI2: Alzheimer's disease neuroimaging initiative 2.

Table 2. Performance results of the MyGNG in each of the eight scenarios studied (MCI-AD classification task).

Scenario						
With AGE	Scaling method	No. PCs	Accu	Spec	Sens	AUC
No	Standard	3	0.909	0.84	0.9391	0.9629
No	Standard	4	0.897	0.7867	0.9449	0.8983
No	Robust	3	0.903	**0.8867**	0.9101	0.9521
No	Robust	4	**0.9131**	0.7867	0.9681	0.9228
Yes	Standard	3	0.9071	0.84	0.9362	**0.966**
Yes	Standard	4	0.9051	0.86	0.9246	0.9389
Yes	Robust	3	0.903	0.84	0.9304	0.9322
Yes	Robust	4	**0.9131**	0.7867	**0.9681**	0.9379

MyGNG: modular hybrid growing neural gas; MCI: mild cognitive impairment; AD: Alzheimer's disease; Accu: accuracy; AUC: area under the curve; PC: principal component; Sens: sensitivity; Spec: specificity. The best values of the different performance metrics are highlighted in bold.

generally considered superior and have been considered the state-of-the-art for the last years.[8] For example, MyGNG was outperformed by Hosseini-Asl et al.[15] and Rashid et al.,[16] which obtained above 0.98 accuracy with convolutional neural network (CNN) or variants. On the other hand, similar or worse results were reported in Basaia et al.,[13] Song et al.[18] and Urooj et al.,[50] especially sensitivity ones as low as 0.68.

CN versus MCI versus AD

Except for the feature selection, the same preprocessing techniques used in the binary dataset were applied to the multiclass one.

This feature selection process began with the ranking of the features with the FCBF method and keeping those with the highest FCBF score for this classification task. This is equivalent to what was done for the MCI-AD task and was described in the "MCI versus AD" subsection. The features ranked by FCBF were in this order, from greater to lesser significance: ["ABETA," "MMDATE," "AGE," "MMDAY," "CLOCKTIME," "MMMONTH," "MMFLOOR," ''NPIG," "MMHOSPIT," "MMYEAR," "MMREAD"]. That is, a cerebrospinal fluid (CSF) value that measures the amount of the amyloid beta (Abeta) biomarker, seven subscores of the

Table 3. Comparison with works from other authors that used ADNI data and dealt with the MCI-AD or the CN-MCI-AD classification tasks.

Task	Works	Neural method	Features	Subjects	Accu	Spec	Sens	AUC
MCI-AD	Hosseini-Asl et al.[15]	3D-DSA-CNN	MRI	70 MCI, 70 AD	1.0	1.0	1.0	
	Basaia et al.[13]	3D-CNN	MRI	253 EMCI, 510 LMCI,	0.86	0.84	0.88	
				294 AD				
	Song et al.[18]	3D-CNN	MRI, PET	160 MCI, 95 AD	0.85	0.95	0.68	
	Urooj et al.[50]	SaDE-WNN	MRI	304 MCI, 258 AD	0.94	0.98	0.86	0.97
	Rashid et al.[16]	Biceph-net (CNN-based)	MRI	N/A for ADNI alone	0.98			
	Current work	MyGNG	NT	345 MCI, 150 AD	0.91	0.84	0.94	0.97
CN-MCI-AD	Zhou et al.[56]	DNN; SVM; SAE+SVM	MRI, PET, SNP	226 CN, 157 sMCI,	0.65			
				389 MCI, 205 pMCI, 190 AD				
	Esmaeilzadeh et al.[57]	3D-CNN	MRI	230 CN, 411 MCI, 200 AD	0.61			
	Basheera and Sai[58]	CNN	MRI	117 CN, 112 MCI, 120 AD	0.87	0.87	0.9	0.89
	Cohen et al.[59]	1D-CNN; ANN	qMRI, APOE,	1299 CN, 1683 MCI, 794 AD	0.88			
			demographic,					
			NT, CSF					
	Sarraf et al.[60]	MCADNNet; DeepAD	sMRI, fMRI	183 CN, 905 MCI, 263 AD	1.0			
	Chen et al.[61]	CNN+iterated RF;	MRI	60 CN, 97 MCI, 43 AD	0.89	0.89	0.89	
		CNN; CNN+SVM;						
		CNN+kNN; CNN+RF						
	Rashid et al.[16]	Biceph-net (CNN-based)	MRI	N/A for ADNI alone	0.97			
	Sharma et al.[62]	CNN+ensemble	sMRI, PET	200 CN, 200 MCI, 200 AD	0.97	0.97	1.0	
		RVFL+RVFL;						
		CNN+RVFL; CNN						
	Khan et al.[63]	VGG	MRI	75 CN, 75 EMCI, 80 LMCI, 85 AD	0.99			
	Current work	MyGNG	qMRI, CSF, NT	229 CN, 402 MCI, 188 AD	0.86	0.8	0.9	0.86

ADNI: Alzheimer's disease neuroimaging initiative; MCI: mild cognitive impairment; CN: cognitively normal; AD: Alzheimer's disease; 3D: three-dimensional; CNN: convolutional neural network; Accu: accuracy; APOE: apolipoprotein E; AUC: area under the curve; CSF: cerebrospinal fluid; DSA: deeply supervised adaptable; fMRI: functional magnetic resonance imaging; MyGNG: modular hybrid growing neural gas; NT: neuropsychological tests; PET: positron emission tomography; pMCI: progressive mild cognitive impairment; qMRI: quantitative magnetic resonance imaging; RVFL: random vector functional link; SaDE-WNN: self-adaptive differential evolution wavelet neural network; SAE: stacked auto-encoder; Sens: sensitivity; sMCI: stable mild cognitive impairment; sMRI: structural magnetic resonance imaging; SNP: single nucleotide polymorphism; Spec: specificity; EMCI: early mild cognitive impairment; LMCI: late mild cognitive impairment; RF: random forest; SVM; support vector machine; kNN: k-nearest neighbor.

MMSE test, an item of the clock drawing test, a subscore of the neuropsychiatric inventory scale, and the age of the patient. In this ranking, ABETA had an FCBF score of 0.4, whereas the second one, MMDATE, 0.15, and the rest, lower than 0.13.

A refining process of this feature set was carried out in order to reduce its size while trying to improve its quality for the multiclass task, which implied selecting other features. This refining process was iterative, one feature at a time, and its goal was to find a new set of features that ensured higher inter-class and lower intra-class distances than the set of features already marked as optimal, starting from the one based on the FCBF ranking. The final set, which was subsequently used to create the input vector of our neural computing system, should be the one where the samples from the same class remain near whereas higher distances exist between those from different classes. This extra process was carried out because of the higher complexity of the task due to being multiclass.

For this refining process, several information on the features were taken into account. Before adding a feature from the available ones, its biological relevance was analyzed according to AD-related clinical bibliography,[51,52] and its adequacy for the diagnostic problem addressed by means of analyzing descriptive statistics (mean, standard deviation, and interval), and clustering quality metrics (Silhouette, Davies–Bouldin, and Calinski–Harabasz scores).[53–55] Conversely, for the discarding of a feature, only the latter two sources were used. After the refining process, the optimal feature vector for this classification task is obtained.

The final set of features comprised, Table 4: a demographic data, a quantitative neuroimaging measurement that measures the volume of the ventricles, a CSF value that measures the quantity of Abeta, and the main scores of two neuropsychological tests (one of them and a subscore of the other were already used for the MCI-AD classification task). It should be noted that some of these features, such as the volume of the ventricles and the final score of the FAQ test, did not appear in the feature set obtained from the FCBF ranking as they were selected afterward during the iterative refining process. Also, all the subscores of the MMSE test were substituted with the total score. The AD-related bibliography was used to confirm if this final set of features was biologically relevant and could be extrapolated to other non-ADNI MCI and AD populations.

The same best combinations of hyperparameters of the MyGNG found in the "MCI versus AD" section, one when 4 PCs were used and another one when 3 PCs were used, were now tested in the CN-MCI-AD classification task. In Table 5, the performance results of the MyGNG in each of the eight scenarios that were studied are shown.

Similar to what was done in the "MCI versus AD" subsection, some works from other researchers that tackled the CN-MCI-AD multiclass classification task and used data from ADNI were selected from the existing literature to

Table 4. Characteristics of the subjects: The five attributes used by the model as input (AGE was used only in half of the scenarios).

	AD	MCI	CN
Number of subjects	188	402	229
AGE: mean	75.26	74.76	75.87
AGE: StD	7.51	7.35	5.01
VENTRICLES: mean	51290.72	45076.82	35785.35
VENTRICLES: StD	26713.07	24276.12	20288.15
ABETA: mean	607.64	741.98	1234.71
ABETA: StD	274.38	345.5	453.74
FAQTOTAL: mean	13.14	3.85	0.14
FAQTOTAL: StD	6.82	4.48	0.6
MMSCORE: mean	23.28	27.01	29.11
MMSCORE: StD	2.03	1.78	1

AD: Alzheimer's disease; MCI: mild cognitive impairment; CN: cognitively normal; StD: standard deviation.

Table 5. Performance results of the MyGNG in each of the eight scenarios studied (CN-MCI-AD classification task).

Scenario						
With AGE	Scaling method	No. PCs	Accu	Spec	Sens	AUC
No	Standard	3	0.8446	0.7587	0.8794	0.8117
No	Standard	4	0.8142	0.7222	0.8611	0.757
No	Robust	3	0.8366	0.7734	0.8867	0.81
No	Robust	4	0.8549	0.7843	0.8922	0.8188
Yes	Standard	3	0.8319	0.7833	0.8916	0.8304
Yes	Standard	4	**0.863**	**0.8036**	**0.9018**	**0.8576**
Yes	Robust	3	0.834	0.7646	0.8823	0.812
Yes	Robust	4	0.859	0.7884	0.8942	0.8311

MyGNG: modular hybrid growing neural gas; CN: cognitively normal; MCI: mild cognitive impairment; AD: Alzheimer's disease; Accu: accuracy; AUC: area under the curve; PC: principal component; Sens: sensitivity; Spec: specificity.
The best values of the different performance metrics are highlighted in bold.

Table 6. Comparison of the MyGNG and several popular neural and non-neural ML methods (MCI-AD classification task).

Method	Accu	Spec	Sens	AUC	CUI+	CUI−
MLP	0.8343	0.8986	0.6867	0.7925	0.5298	0.7927
RF	0.9172	0.9623	0.8133	0.8876	0.7364	0.8881
DT	0.899	0.9391	0.8067	0.8728	0.6888	0.8622
NB	0.9152	0.9391	0.86	0.8994	0.7406	0.8823
SVM	**0.9293**	**0.9739**	0.8267	0.9	0.771	**0.9043**
MyGNG	0.907	0.84	**0.9362**	**0.966**	**0.8722**	0.721

Accu: accuracy; AUC: area under the curve; CUI: clinical utility index; DT: decision tree; MLP: multilayer perceptron; MyGNG: modular hybrid growing neural gas; NB: Naïve Bayes; RF: random forest; Sens: sensitivity; Spec: specificity; SVM: support vector machine; ML: machine learning.
The best values of the different performance metrics are highlighted in bold.

Table 7. Comparison of the MyGNG and several popular neural and non-neural ML methods (CN-MCI-AD classification task).

Method	Accu	Spec	Sens	AUC	CUI+	CUI−
MLP	0.6874	0.8437	0.6874	0.749	0.4755	0.7126
RF	0.8291	**0.9243**	0.8291	**0.8675**	0.7014	**0.8461**
DT	0.7558	0.9078	0.7558	0.8318	0.6083	0.8004
NB	0.8035	0.9017	0.8035	0.8532	0.6469	0.8135
SVM	0.8046	0.9023	0.8046	0.846	0.6481	0.8143
MyGNG	**0.863**	0.8036	**0.9018**	0.8576	**0.7442**	0.7305

ML: machine learning; Accu: accuracy; AUC: area under the curve; CUI: clinical utility index; DT: decision tree; MLP: multilayer perceptron; MyGNG: modular hybrid growing neural gas; NB: Naïve Bayes; RF: random forest; Sens: sensitivity; Spec: specificity; SVM: support vector machine; MCI: mild cognitive impairment; CN: cognitively normal; AD: Alzheimer's disease.
The best values of the different performance metrics are highlighted in bold.

qualitatively compare our method with, Table 3. Generally and as expected, performance results are lower than in the binary task. MyGNG yielded equivalent performance results, sometimes higher. Unlike in the binary task, our system made use of, apart from neuropsychological tests as in the binary task, quantitative neuroimaging and CSF. Those approaches that outperformed MyGNG varied from CNN variants (such as the VGG in Khan et al.[63]) to non-monolithic approaches, such as modular methods that combined a CNN whether with an ensemble of random forest (RF)[61] or with an ensemble of random vector functional link (RVFL) networks plus a following module made from a single RVFL.[62]

Comparative studies with ML models

For both classification tasks our ANN-based method, MyGNG, was compared with several popular supervised ML models: a decision tree (DT) (flowchart-like structure; easier to interpret than ANN),[64] a Naïve Bayes (NB) classifier (based on applying the Bayes' theorem and assuming that the features are strongly independent given class),[65] an RF (an ensemble of DT, each trained with a random subset of features; the class that is returned is the one chosen by most DT),[64] a support vector machine (SVM) (builds an hyperplane usually in a high-dimensional space to separate classes; using certain kernel functions allow separation of non-linear data),[66] and a multilayer perceptron (MLP) (a feedforward ANN able to separate non-linear data, unlike the perceptron).[67]

These ML models were implemented with "scikit-learn" and "Keras," popular ML and DL Python modules, respectively.[38,68] Best results were achieved by models with the following combinations of hyperparameters. For DT,

Pruning = at least two instances in leaves; at least five instances in internal nodes; maximum depth = 100; Splitting: Stop splitting when majority reaches 95% (classification only); Binary trees: Yes for NB, scikit-learn's default values. For RF, number of trees = 10; maximal number of considered features = unlimited; replicable training = No; maximal tree depth = unlimited; stop splitting nodes with maximum instances = 5. For SVM, $C = 1.0$, $\varepsilon = 0.1$; Kernel: radial basis function; $\exp(-auto|x - y|^2)$; numerical tolerance: 0.001; iteration limit: 100. For MLP, hidden neurons = (16, 8); activation function = "relu"; solver = "rmsprop."

In Table 6, we compared the MyGNG performance results with those obtained by other popular ML methods in the scenario which was found optimal for the MyGNG: with AGE, standard scaling, and 3 PCs.

A similar comparison can be found in Table 7 for the CN-MCI-AD classification task. In this case, the scenario that was considered optimal for the MyGNG was: with AGE, standard scaling and 4 PCs.

Discussion

Throughout the "MCI versus AD" and "CN versus MCI versus AD" subsections it has been shown that our intelligent system based on the MyGNG is quite competent for both classification tasks related to the early diagnosis of AD, MCI-AD and CN-MCI-AD.

According to the eight different scenarios based on the three questions that we wanted to reply, the MyGNG considers beneficial data that have been scaled in the standard way, data that have been projected with 4 PCs, and adding AGE to the feature set. The latter occurred for the rest of ML algorithms
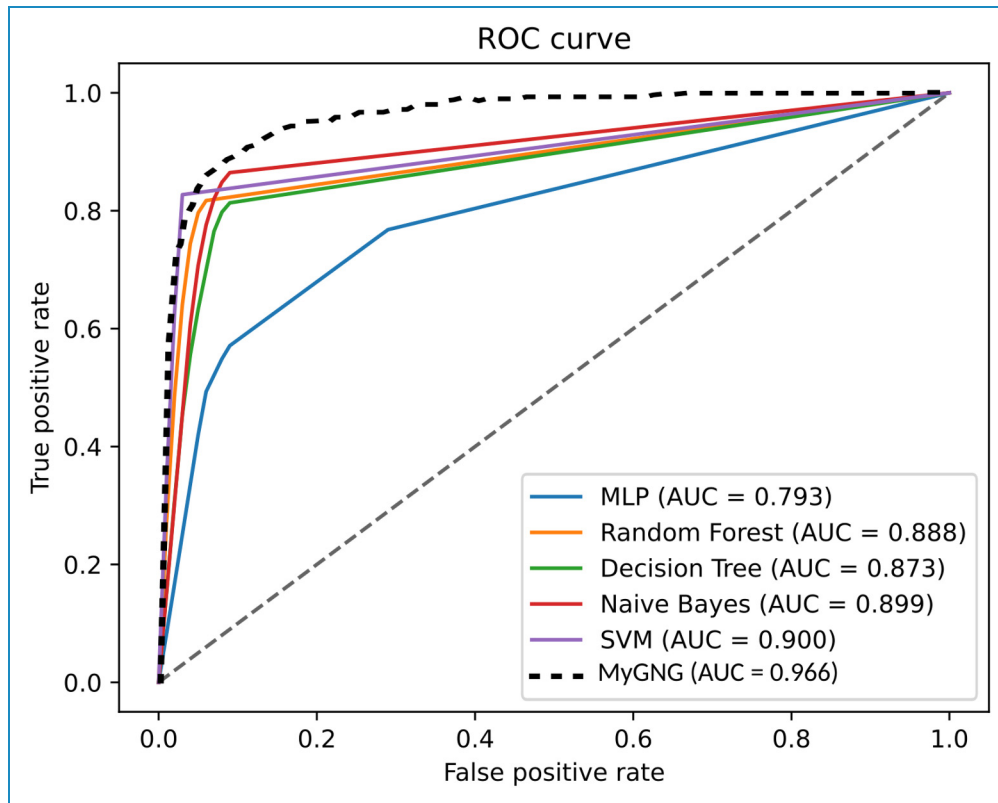
Figure 3. ROC curves of the MyGNG and the ML algorithms in the MCI-AD classification task. ROC: receiver operating characteristic; MyGNG: modular hybrid growing neural gas; ML: machine learning; MCI: mild cognitive impairment; AD: Alzheimer's disease.

as well, hence demonstrating the usefulness of also using this demographic information that is considered an AD risk factor.

Despite the fact that features obtained with neuroimaging techniques, mainly MRI, and CNN models, which are considered the state-of-the-art and prevalent combination, were used in all but one of the works found for MCI-AD and that used ADNI data,[15,13,18,50,16] our MyGNG and neuropsychological tests combination performed better than some of them. Also, it required not only a fraction of the training time and computational power but also several orders of magnitude lower number of hyperparameters to tune. Taking into account the most reliable performance metric robust to class unbalanced datasets of those used in this work (i.e. AUC), the MyGNG was only outperformed by one of the ML models in CN-MCI-AD, Table 7, whereas it yielded the best values in MCI-AD, Table 6. In CN-MCI-AD, the winner ML model was an RF, a type of ensemble. Ensemble-based systems are generally considered more advantageous and powerful than single-expert systems,[69] as is the MyGNG. In addition to this, and according to our results, this gain existed but only occurred in the multiclass problem, which is more complex. In both classification tasks, the MyGNG yielded the highest CUI+ and sensitivity values. The latter indicates that it preferred to classify a subject as

AD, the minority class, which is when more worrisome are the symptoms and more treatment is needed, something that is usually of interest to clinicians and health systems. On the contrary, the accuracy and specificity of the MyGNG were lower, values for the first metric might be explained by the class unbalanced dataset.

Our MyGNG was also compared with other ML models graphically presented by ROC curves, Figures 3 and 4. They showed that the MLP performed poorly in both tasks, whereas our MyGNG outperformed the others in the binary classification task, and all but one in the multiclass case. This is especially noticeable when the false positive rate (FPR) has higher values in both tasks and for almost any FPR values in the binary task. The rest of the ML algorithms behaved similarly to each other.

Regarding the clinical relevance of our model, it is focused on primary care, albeit it can also be utilized in specialized one. The criteria that have been used are fast and non-expensive, even more so when in some cases the subscores of neuropsychological scales and not the total scores are required. As less time is needed for each patient, their quality of life might improve. On the other hand, performance results have been reported with the CUI metric, which measures the real clinical utility of diagnostic criteria so some authors consider it useful and
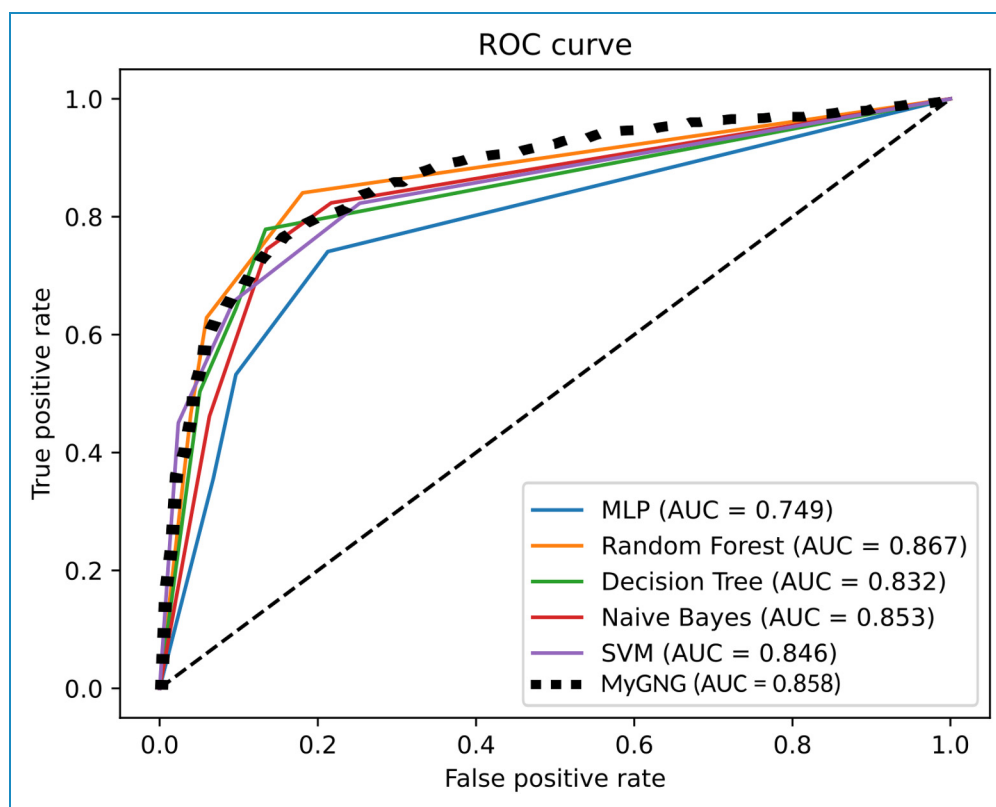
Figure 4. ROC curves of the MyGNG and the ML algorithms in the CN–MCI-AD classification task. ROC: receiver operating characteristic; MyGNG: modular hybrid growing neural gas; ML: machine learning; CN: cognitively normal; MCI: mild cognitive impairment; AD: Alzheimer's disease.

relevant for clinicians.[39,10] According to the interpretation in Mitchell,[39] MyGNG yielded CUI+ and CUI− values considered "good" in the multiclass classification task, increasing to "excellent" for the CUI+ in the binary task. Therefore, the computational solution for AD diagnosis presented in this work has a clinical utility good enough to make it adequate as a translational medicine product. Finally, it is planned to integrate the MyGNG in eHealth systems, such as EDEVITALZH,[70] which will enable us to make diagnoses anywhere and at any time, and therefore lead us toward universal diagnosis. Furthermore, this integration will bring along some valuable information regarding the generalization capabilities of our MyGNG-based system.

This study has potential limitations. Regarding data, ADNI was the source of all the data we used, which might imply two shortcomings: a possibly reduced generalization to older populations from other regions, and the limitation to the cohorts/classes, modalities, and clinical criteria available in this database. Examples of these three limitations are: patients in ADNI have not been categorized within the AD continuum by means of Amyloid-Tau-Neurodegeneration (ATN) profiles (however, it could be possible with some of the criteria already available in ADNI[71]); there is no optical coherence tomography angiography data; and several neuropsychological scales are unavailable). Unfortunately, no solution to these limitations is possible within ADNI unless ADNI procedures change, so other databases, probably private, would be required to surpass these shortcomings. About the method, MyGNG, although it can also be used in specialized attention, has been primarily focused on primary care, so it is currently unable to work with neuroimaging data unless they are in a quantitative format. Finally, our comparison with other authors' works was limited, so a broader one will be of interest to researchers, especially if approaches based on non-neural methods are also included.

Conclusions

In this work, we have really improved the first proposal of the ontogenic ANN MyGNG, by just changing the supervised module for a simpler one, nothing more than a perceptron algorithm scheme. We have analyzed the behavior of the improved MyGNG in two classification tasks related to the early diagnosis of AD: MCI-AD and CN-MCI-AD. According to our results, this MyGNG proved to be a better computational solution than the other ML algorithms that were compared with and was only slightly outperformed by an ensemble method, an RF, in the multiclass task regarding some of the used performance measurements.

Additionally, in both tasks, qualitative comparisons with proposals from other authors delivered surprising results, as most of them were DL-based and made use of data from neuroimaging techniques, the state-of-the-art. In MCI-AD, some of those works were outperformed by our MyGNG combined with only six features derived from three neuropsychological tests. Similarly happened in CN-MCI-AD with demographic data, a quantitative neuroimaging measurement, a CSF value, and two neuropsychological tests.

The main contributions of this work can be grouped into clinical and computational. In the first group, there are several. Our dataset is built with features obtained with non-invasive modalities, that are fast to collect. Another increase in speed in the data-gathering process is obtained thanks to the usage of subscales instead of the total score of the neuropsychological scales. A major advantage of our system is its low complexity, so it is convenient for both primary and specialized care and can be used not only in hospitals and medical sites but also in sociosanitary institutions. In the second group, our system is based on a new neural architecture, the MyGNG. This neural architecture is an ontogenic ANN so it is more adaptive to the space of the problem because it is also able to modify its structure, adjusting more and better to that space. Furthermore, a faster learning is enabled by its hybrid nature. On the other hand, the MyGNG has outperformed several DL approaches in the same classification tasks. Non-deep solutions as the one presented in this work have several times fewer parameters to configure, train a lot faster, and do not require expensive and complex hardware for the training process to be done in a reasonable time. Moreover, due to its modular design, it is possible for both modules in the MyGNG to learn dynamically. Finally, as with other methods, MyGNG can be integrated into eHealth systems, allowing its online use.

Regarding future works, the MyGNG tackling other classification tasks will be worthwhile. Further analysis of the features and preprocessing techniques may be helpful in these classification tasks, and probably mandatory with others. More complex ontogenetic neural architectures may arise, which will be more powerful and might perform optimally in these tasks. Other technical advances can also be incorporated such as better validation frameworks and faster computing, among others.

ORCID iDs: Ylermi Cabrera-León https://orcid.org/0000-0001-5709-2274
Pablo Fernández-López https://orcid.org/0000-0002-2135-6095
Konrad Kluwak https://orcid.org/0000-0003-0983-2228

Notes

1. More up-to-date information on ADNI can be found at adni.loni.usc.edu.
2. In this work, s_i indicates the neuron with the ith minimum Euclidean distance between the input vector and the neuron's weights vector.

References

1. Gauthier S, Webster C, Servaes S, Morais JA and Rosa-Neto P. World Alzheimer Report 2022 – Life after diagnosis: Navigating treatment, care and support. Technical report, Alzheimer's Disease International, 2022.
2. Reitz C and Mayeux R. Alzheimer disease: Epidemiology, diagnostic criteria, risk factors and biomarkers. *Biochem Pharmacol* 2014; 88: 640–651.
3. World Health Organization. *Risk reduction of cognitive decline and dementia: WHO guidelines*. Geneva, Switzerland: World Health Organization, 2019.
4. Zhu CW and Sano M. Economic considerations in the management of Alzheimer's disease. *Clin Interv Aging* 2006; 1: 143–154.
5. Alzheimer's Association. Alzheimer's disease facts and figures. *Alzheimers Dement* 2023; 19: 1598–1695.
6. Blennow K and Zetterberg H. Biomarkers for Alzheimer's disease: Current status and prospects for the future. *J Intern Med* 2018; 284: 643–663.
7. González-Sánchez M, Bartolome F, Antequera D, et al. Decreased salivary lactoferrin levels are specific to Alzheimer's disease. *EBioMedicine* 2020; 57: 102834.
8. Cabrera-León Y, García Báez P, Fernández-López P, et al. Neural computation-based methods for the early diagnosis and prognosis of Alzheimer's disease not using neuroimaging biomarkers: A systematic review. *J Alzheimers Dis* 2024; 98: 793–823.
9. Janoutová J, Šerý O, Hosák L, et al. Is mild cognitive impairment a precursor of Alzheimer's disease? Short review. *Cent Eur J Public Health* 2015; 23: 365–367.
10. Mitchell AJ and Shiri-Feshki M. Rate of progression of mild cognitive impairment to dementia – meta-analysis of 41 robust inception cohort studies. *Acta Psychiatr Scand* 2009; 119: 252–265.
11. Erkinjuntti T, Østbye T, Steenhuis R, et al. The effect of different diagnostic criteria on the prevalence of dementia. *New England J Med* 1997; 337: 1667–1674.
12. Wancata J, Börjesson-Hanson A, Östling S, et al. Diagnostic criteria influence dementia prevalence. *Am J Geriatr Psychiatry* 2007; 15: 1034–1045.
13. Basaia S, Agosta F, Wagner L, et al. Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *Alzheimer's Dement: Transl Res Clin Interv* 2019; 5: 974–986.
14. Ebrahimi-Ghahnavieh A, Luo S and Chiong R. Transfer learning for Alzheimer's disease detection on MRI images. In *2019 IEEE international conference on industry 4.0, Artificial intelligence, and communications technology (IAICT)*, Bali, Indonesia, 01–03 July 2019, pp.133–138. New York City, NY, USA: IEEE.
15. Hosseini-Asl E, Ghazal M, Mohmoud A, et al. Alzheimer's disease diagnostics by a 3D deeply supervised adaptable convolutional network. *Front Biosci* 2018; 23: 584–596.
16. Rashid AH, Gupta A, Gupta J, et al. Biceph-net: A robust and lightweight framework for the diagnosis of Alzheimer's disease using 2D-MRI scans and deep similarity learning. *IEEE J Biomed Health Inform* 2022; 27: 1205–1213.
17. Santander-Cruz Y, Salazar-Colores S, Paredes-García WJ, et al. Semantic feature extraction using SBERT for dementia detection. *Brain Sci* 2022; 12: 1–18.
18. Song J, Zheng J, Li P, et al. An effective multimodal image fusion method using MRI and PET for Alzheimer's disease diagnosis. *Front Digit Health* 2021; 3: 637386.
19. Cabrera-León Y, Báez PG, Ruiz-Alzola J, et al. Classification of mild cognitive impairment stages using machine learning methods. In: *2018 IEEE 22nd international conference on intelligent engineering systems (INES)* Las Palmas de Gran Canaria, Spain, June 2018, pp.67–72. New York City, NY, USA: IEEE.
20. Kruthika KR, Rajeswari and Maheshappa HD. Multistage classifier-based approach for Alzheimer's disease prediction and retrieval. *Inform Med Unlocked* 2019; 14: 34–42.
21. Ramzan F, Khan MUG, Rehmat A, et al. A deep learning approach for automated diagnosis and multi-class classification of Alzheimer's disease stages using resting-state fMRI and residual neural networks. *J Med Syst* 2019; 44: 37.
22. Cui R, Liu M and Li G. Longitudinal analysis for Alzheimer's disease diagnosis using RNN. In: *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, Washington, DC, USA, 04–07 April 2018, pp.1398–1401. New York City, NY, USA: IEEE.
23. Tabarestani S, Aghili M, Eslami M, et al. A distributed multitask multimodal approach for the prediction of Alzheimer's disease in a longitudinal study. *NeuroImage* 2020; 206: 116317.
24. Manzak D, Çetinel G and Manzak A. Automated classification of Alzheimer's disease using deep neural network (DNN) by random forest feature elimination. In: *2019 14th international conference on computer science education (ICCSE)*, Toronto, ON, Canada, 19–21 August 2019, pp.1050–1053. New York City, NY, USA: IEEE.
25. Cabrera-León Y, Báez PG, Fernández-López P, et al. Study on mild cognitive impairment and Alzheimer's disease classification using a new ontogenic neural architecture, the supervised reconfigurable growing neural gas. In: *2023 annual modeling and simulation conference (ANNSIM 2023)* Mohawk College, ON, Canada, May 2023, pp.425–436. Los Alamitos, CA, USA: IEEE Computer Society.

26. Sabbaghi N, Sheikhani A, Noroozian M, et al. Interval-based features of auditory ERPs for diagnosis of early Alzheimer's disease. *Alzheimers Dement (Amst)* 2021; 13: e12191.

27. Sosa-Marrero A, Cabrera-León Y, Fernández-López P, et al. Detection of Alzheimer's disease versus mild cognitive impairment using a new modular hybrid neural network. In: Rojas I, Joya G and Catala A (eds) *Advances in computational intelligence* (*Lecture notes in computer science*). Cham: Springer International Publishing, 2021, pp. 223–235.

28. Suárez-Araujo CP, García Báez P, Cabrera-León Y, et al. A real-time clinical decision support system, for mild cognitive impairment detection, based on a hybrid neural architecture. *Comput Math Methods Med* 2021; 2021: 1–9.

29. Pellegrini E, Ballerini L, Valdes Hernandez MDC, et al. Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: A systematic review. *Alzheimer's Dement: Diagnosis, Assess Disease Monit* 2018; 10: 519–535.

30. Yao X, Yan J, Ginda M, et al. Mapping longitudinal scientific progress, collaboration and impact of the Alzheimer's disease neuroimaging initiative. *PLoS ONE* 2017; 12: e0186095.

31. Fritzke B. A growing neural gas network learns topologies. *Adv Neural Inf Process Syst* 1995; 0: 625–632.

32. Fiesler E. Comparative bibliography of ontogenic neural networks. In: Marinaro M and Morasso PG (eds) *ICANN '94*. London: Springer, pp.793–796.

33. Fritzke B. Unsupervised ontogenic networks. In: *Handbook of neural computation*. Computational Intelligence Library, Boca Ratón, FL, USA: CRC Press, 1997, p.16.

34. Rosenblatt F. Principles of neurodynamics. Perceptrons and the theory of brain mechanisms. Technical Report VG-1196-G-8, Cornell Aeronautical Lab Inc., Buffalo, New York, 1961.

35. Widrow B and Lehr MA. 30 years of adaptive neural networks: Perceptron, madaline, and backpropagation. *Proc IEEE* 1990; 78: 1415–1442.

36. Kamimura R. Information enhancement for interpreting competitive learning. *Int J Gen Syst* 2010; 39: 705–728.

37. Mendes CAT, Gattass M and Lopes H. FGNG: A fast multidimensional growing neural gas implementation. *Neurocomputing* 2014; 128: 328–340.

38. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res* 2011; 12: 2825–2830.

39. Mitchell AJ. A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. *J Psychiatr Res* 2009; 43: 411–431.

40. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett* 2006; 27: 861–874.

41. Aggarwal CC and Reddy CK (eds) *Data clustering: Algorithms and applications*. 1st ed. New York: Chapman and Hall/CRC, 2016.

42. Chen T and Guestrin C. XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining – KDD '16*, San Francisco, California, USA, 13–17 August 2016, pp.785–794. New York NY, United States: Association for Computing Machinery.

43. Yu L and Liu H. Feature selection for high-dimensional data: A fast correlation-based filter solution. In: *Proceedings of the twentieth international conference on machine learning (ICML 2003)*, Washington, DC USA, 21–24 August 2003, pp.856–863. Cambridge, MA, USA: AAAI Press.

44. Folstein MF, Folstein SE and McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975; 12: 189–198.

45. Rosen WG, Mohs RC and Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry* 1984; 141: 1356–1364.

46. Pfeffer RI, Kurosaki TT, Harrah CH, et al. Measurement of functional activities in older adults in the community. *J Gerontol* 1982; 37: 323–329.

47. García Báez P, Suárez Araujo CP, Viadero CF, et al. Automatic prognostic determination and evolution of cognitive decline using artificial neural networks. In: Yin H, Tino P, Corchado E, Byrne W and Yao X (eds) *Intelligent data engineering and automated learning – IDEAL 2007* (*Lecture notes in computer science*, vol. 4881). Berlin: Springer, 2007, pp. 898–907.

48. Jolliffe IT and Cadima J. Principal component analysis: A review and recent developments. *Philos Trans R Soc A: Math Phys Eng Sci* 2016; 374: 20150202.

49. Goldberger J, Roweis S, Hinton G, et al. Neighbourhood components analysis. In: *NIPS'04: Proceedings of the 17th international conference on neural information processing systems*, Vancouver, British Columbia, Canada, 1 December 2004, pp.513–520. Cambridge MA, USA: MIT Press.

50. Urooj S, Singh SP, Malibari A, et al. Early detection of Alzheimer's disease using polar harmonic transforms and optimized wavelet neural network. *Appl Sci* 2021; 11: 1574.

51. Budelier MM and Bateman RJ. Biomarkers of Alzheimer disease. *J Appl Lab Med* 2020; 5: 194–208.

52. Gunes S, Aizawa Y, Sugashi T, et al. Biomarkers for Alzheimer's disease in the current state: A narrative review. *Int J Mol Sci* 2022; 23: 4962.

53. da Silva SF, Brandoli B, Eler DM, et al. Silhouette-based feature selection for classification of medical images. In: *2010 IEEE 23rd international symposium on computer-based medical systems (CBMS)*, Perth, Australia, 12–15 October 2010, pp.315–320. Los Alamitos, CA, USA: IEEE Computer Society.

54. Gutoski M, Ribeiro M, Aquino NMR, et al. Feature selection using differential evolution for unsupervised image clustering. In: Rutkowski L, Scherer R, Korytkowski M, et al. (eds) *Artificial intelligence and soft computing. ICAISC 2018*, Zakopane, Poland, vol. 10841, 2018. Cham: Springer.

55. McCrory M and Thomas SA. Cluster metric sensitivity to irrelevant features. *arXiv computer science, machine learning*, arXiv:2402.12008.

56. Zhou T, Thung K-H, Zhu X, et al. Feature learning and fusion of multimodality neuroimaging and genetic data for multi-status dementia diagnosis. In: Wang Q, Shi Y, Suk H-I and Suzuki K (eds) *Machine learning in medical imaging*, vol. 10541. Cham: Springer International Publishing, 2017, pp.132–140.

57. Esmaeilzadeh S, Belivanis DI, Pohl KM, et al. End-to-end Alzheimer's disease diagnosis and biomarker identification. In: Shi Y, Suk H-I and Liu M (eds) *Machine learning in medical imaging*, vol. 11046. Cham: Springer International Publishing, 2018, pp.337–345.

58. Basheera S and Sai Ram MS. Convolution neural network–based Alzheimer's disease classification using hybrid enhanced independent component analysis based segmented gray matter of T2 weighted magnetic resonance imaging with clinical valuation. *Ann Transl Med* 2019; 10: 765.

59. Cohen DS, Carpenter KA, Jarrell JT, et al. Deep learning-based classification of multi-categorical Alzheimer's disease data. *Curr Neurobiol* 2019; 10: 141–147.

60. Sarraf S, Desouza DD, Anderson JAE, et al. Semantic feature extraction using SBERT for dementia detection. *IEEE Access* 2019; 7: 155584–155600.

61. Chen X, Tang M, Liu A, et al. Diagnostic accuracy study of automated stratification of Alzheimer's disease and mild cognitive impairment via deep learning based on MRI. *Ann Transl Med* 2022; 10: 765.

62. Sharma R, Goel T, Tanveer M, et al. Conv-ERVFL: Convolutional neural network based ensemble RVFL classifier for Alzheimer's disease diagnosis. *IEEE J Biomed Health Inform* 2022; 27: 4995–5003.

63. Khan R, Akbar S, Mehmood A, et al. A transfer learning approach for multiclass classification of Alzheimer's disease using MRI images. *Front Neurosci* 2023; 16: 1050777.

64. Kotsiantis SB. Decision trees: A recent overview. *Artif Intell Rev* 2013; 39: 261–283.

65. Rish I. An empirical study of the naive Bayes classifier. Technical Report RC 22230 (W0111-014), IBM, 2001.

66. Cortes C and Vapnik V. Support-vector networks. *Mach Learn* 1995; 20: 273–297.

67. Haykin SS. *Neural networks and learning machines*. New Jersey, USA: Pearson Prentice Hall, 2009.

68. Chollet F, & others. 2015. Keras. GitHub. Retrieved from https://keras.io/

69. Polikar R. Ensemble based systems in decision making. *IEEE Circuit Syst Mag* 2006; 6: 21–45.

70. Suárez-Araujo CP, Pino P, Ángel M, et al. EDEVITALZH: Predictive, preventive, participatory and personalized e-Health platform to assist in the geriatrics and neurology clinical scopes. In: *International conference on computer aided systems theory (EUROCAST 2011)*, Las Palmas de Gran Canaria, Spain, 6–11 February 2011, pp.264–271. Berlin, Heidelberg: Springer.

71. Marcolini S, Mondragón JD, Dominguez-Vega ZT, et al. Clinical variables contributing to the identification of biologically defined subgroups within cognitively unimpaired and mild cognitive impairment individuals. *Eur J Neurol* 2024; 31: 1468–1331.