ORIGINAL ARTICLE

GA-MADRID: design and validation of a machine learning tool for the diagnosis of Alzheimer's disease and frontotemporal dementia using genetic algorithms

Fernando García-Gutierrez[1] · Josefa Díaz-Álvarez[2] · Jordi A. Matias-Guiu[1] · Vanesa Pytel[1] · Jorge Matías-Guiu[1] ·
María Nieves Cabrera-Martín[1] · José L. Ayala[3]

Abstract

Artificial Intelligence aids early diagnosis and development of new treatments, which is key to slow down the progress of the diseases, which to date have no cure. The patients' evaluation is carried out through diagnostic techniques such as clinical assessments neuroimaging techniques, which provide high-dimensionality data. In this work, a computational tool is presented that deals with the data provided by the clinical diagnostic techniques. This is a Python-based framework implemented with a modular design and fully extendable. It integrates (i) data processing and management of missing values and outliers; (ii) implementation of an evolutionary feature engineering approach, developed as a Python package, called **PyWinEA** using Mono-objective and Multi-objetive Genetic Algorithms (NSGAII); (iii) a module for designing predictive models based on a wide range of machine learning algorithms; (iv) a multiclass decision stage based on evolutionary grammars and Bayesian networks. Developed under the eXplainable Artificial Intelligence and open science perspective, this framework provides promising advances and opens the door to the understanding of neurodegenerative diseases from a data-centric point of view. In this work, we have successfully evaluated the potential of the framework for early and automated diagnosis with neuro-images and neurocognitive assessments from patients with Alzheimer's disease (AD) and frontotemporal dementia (FTD).

Keywords Alzheimer's disease · Frontotemporal dementia · Neurodegenerative diseases · Machine learning · Artificial Intelligence

1 Introduction

Artificial Intelligence (AI) provides innovative solutions to solve complex real-world problems. Machine learning (ML) is one of its most representative branches with the fastest growing. The health sector frequently generates a large volume of highly dimensional data as those produced by neuroimaging techniques, such as magnetic resonance imaging (MRI) and positron emission tomography (PET) [10]; ML algorithms help on providing diagnosis, decisions, or even predictions related to the health status of patients.

✉ Josefa Díaz-Álvarez
mjdiaz@unex.es

Fernando García-Gutierrez
ga.gu.fernando@gmail.com

Jordi A. Matias-Guiu
jordimatiasguiu@hotmail.com

Vanesa Pytel
vanesa.pytel@gmail.com

Jorge Matías-Guiu
matiasguiu@gmail.com

María Nieves Cabrera-Martín
mncabreram@hotmail.com

José L. Ayala
jayala@ucm.es

1   Departments of Neurology, Hospital Clinico San Carlos,
    San Carlos Research Health Institute (IdISSC), Universidad
    Complutense, Madrid, Spain

2   Department of Computer Architecture and Communications,
    Centro Universitario de Mérida, Universidad de
    Extremadura, Mérida, Spain

3   Department of Computer Architecture and Automation,
    Universidad Complutense, Madrid, Spain

Adjusting the hyperparameters of ML algorithms to get the best performance is not a trivial task; it requires expertise [36]. Therefore, ML models need to be endowed with explainability and transparency on the basis of the eXplainable Artificial Intelligence (XAI) paradigm [28], which will generate confidence and reliability in the results. This fact is connected to AI democratization [35] and the open science perspective, where sharing and collaborating are two essential objectives.

The interest of the scientific and medical community in providing solutions based on AI to enhance and assist in the diagnosis, prevention and/or development of new treatments has increased significantly [45]. Despite the assistance, caution is needed to prevent any unintended though negative consequences that may occur, for instance if some data are not contextualised [9].

Mentioning some scientific literature in this domain, [45] analysed the potential of AI and ML for the medicine field, and identified changes and challenges to reach accurate and comprehensive diagnosis. [1] presented a review of different solutions, approaches and perspectives of AI and ML, especially for the healthcare sector. Authors included a critical vision, where they pointed out some issues to be improved in order to guarantee the privacy and data security, and to enhance accuracy. Recently, [52] provided a review on current computational approaches applied in the spectrum of neurodegenerative diseases.

Focusing on neurodegenerative diseases and *Python*-developed studies, [29] performed a ML-based analysis to perform data-driven diagnosis of dementia and used post-mortem confirmed cases as a gold-standard; [43] implemented a pipeline based on a DeepSymNet architecture to detect the AD progression pattern. Recently, [53] applied the feature engineering to build voice biomarkers and improve the early detection of Parkinson disease. [8] analysed a group of individuals diagnosed with both behavioural and language variants FTD, using a deep learning algorithm. [17] assessed ^{18}F-2-fluoro-2-deoxy-D-glucose positron emission tomography (18F-FDG PET) brain images from Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset and a retrospective independent test set through a convolutional neural network of InceptionV3. [14] tackled the classification of Alzheimer's disease into four classes using 3D Diffusion Tensor Imaging (3D-DTI) processing.

Neurodegenerative diseases include a wide spectrum of disorders with different clinical manifestations and pathological patterns, where an early accurate diagnosis is challenging. AD is one of the most prevalent [20] and causes a progressive and irreversible brain damage that prevents patients from performing daily activities. FTD is the third cause of dementia, particularly the behavioral variant (bvFTD), and its onset occurs at middle-age [20]. In this work, we specifically focus on AD and FTD, although other neurodegenerative diseases could be similarly addressed by our proposal.

The assessment of patients who suffer from neurodegenerative disorders entails the application of neuroimaging techniques, neuropsychological tests and the clinical histories [23]. Neuropsychological tests assess the cognitive function affected by AD and FTD. Among the neuroimaging techniques, 18F-FDG PET is a minimally Invasive technique. 18F-FDG PET gives a map of brain coordinates associated to metabolism rates, which measure the alterations of glucose consumption in the brain. The presence of alterations in brain metabolism has proven to be a useful biomarker for early diagnosis of AD and FTD [7, 26, 39]. These techniques provide a large volume of data [10], which require experts to be trained in their analysis and interpretation, but the risk of inaccurate diagnoses is real, especially considering the need of early detection of these disorders [18]. In this context, ML techniques are a reliable alternative for design decision-making models that support specialists in the early diagnosis of the disease, monitoring and designing personalized treatments [45], where accuracy is extremely important.

Last decade, many researchers have demonstrated their potential for supporting decisions-making in the clinical arena [11, 25, 34]. However, to the best of our knowledge, we cannot find any other framework in the literature that targets the fully automated diagnosis of AD and FTD from multiple and heterogeneous data sources. The proposed computational tool embodies all the required steps to deal with the data modelling process.

Thus, it integrates the following functionalities:

1. An automate methodology for dataset preprocessing, including imputation techniques to deal with missing, outliers and categorization of nominal variables.
2. A feature engineering module implemented by means of evolutionary algorithms to extract the most relevant features for the diagnosis.
3. A meta-model based on evolutionary grammars and Bayesian networks (BN) for multi-class classification.
4. A basic visualization tool.
5. Different tools for assessing the results.

This work is structured as follows. Firstly, the framework designed is presented. Subsequently, results of the tests using the proposed AI-based tool are summarized and the following section discusses the results of the test case. Next, the conclusion are presented. Finally, the general methodology is described.
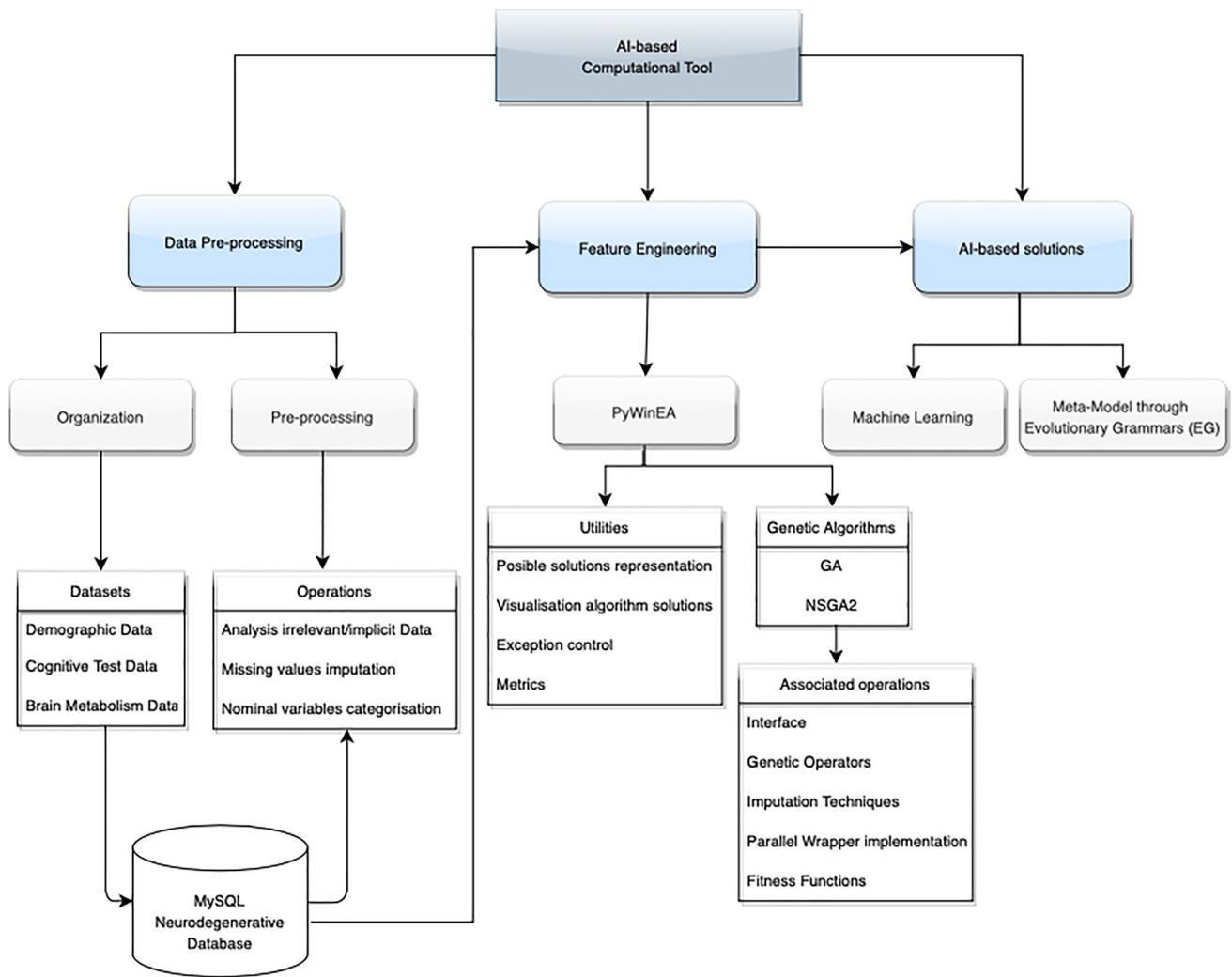
Fig. 1 General scheme of the AI framework proposed, including data pre-processing, feature engineering and IA-based modelling

2 Methods

This is a Python-based framework that makes the data modelling easier to be computed and it is fully extendable thanks to its modular design from the data-driven point of view. According to the general scheme presented in Fig. 1, this Python-based framework provides resources to address data pre-processing, feature selection, a wide set of machine learning models, different AI-based modelling strategies with mono-objective and multi-objective evolutionary algorithms. It also implements a multiclass classification model using EG or Bayesian classifiers. In addition, it provides graphical evaluation tools based on different metrics to asses the results obtained.

Focusing on supplying a fast, robust and reliable AI-based tool, the proposed framework is able to deal with different datasets, including cognitive evaluation, neuroimaging techniques, and the patients' history to help in the diagnosis

of AD and FTD, two neurodegenerative diseases that may present similar symptoms and cognitive and behavioral deficits. Although episodic memory dysfunction is one of the cognitive hallmarks of AD, FTD usually presents also these symptoms. Similarly, behavioral deficits are increasingly recognized symptoms in AD [24, 41]. The management and organisation of data are carried out through a relational database, particularly MySQL. Data are structured in indexed tables that ensure the accessibility, availability and simplifies the data preprocessing. Additionally, relational databases and processes are implemented to easily incorporate new data and guarantee the data integrity.

As aforementioned, this framework manages three types of data: (1) Demographic Data provide variables that describe the sample; (2) Cognitive Test Data contain variables associated with cognitive tests, where each cognitive test provides a rating scale and scores to identify specific kind of cognitive problems and abilities. These tests gather

information about the following cognitive function: memory, visuospatial, executive, attention and language. (3) Brain Metabolism Data include the brain hypometabolism data from the FDG-PET analysis.

Regarding the brain regions, this framework considers two different atlases, the Brodmann's atlas (47 regions) [6] and the Automated Anatomical Labelling (AAL) atlas (90 regions)[54]. Data related to the brain metabolism are divided into qualitative and quantitative. Quantitative data are defined by the number of hypometabolic voxels in a given region. A voxel is a 3D unit of an image, which can be associated with a single value, such as metabolism. Hypometabolic voxels are computed through the voxel-based mapping analysis against a healthy control group. The qualitative data indicate whether a certain area is hypometabolic or not. Although the number of voxels needed to consider a regions as hypometabolic may vary, in this study we selected a threshold of 1 voxel in each region. Therefore, a region was defined as hypometabolic when it has one or more hypometabolic voxels. Although we agree that this is a very limiting threshold, the purpose of this work is to present a parameterizable computing framework, in which this threshold, as many other parameters, can be selected by the expert user in order to meet its clinical goals. The clinical value of the results obtained by the use of our proposed framework is out of the scope of this publication, but is has been already proven in [27].

In order to reduce the effort to reproduce experiments, adapt the implementation to the XAI perspective and gain trust and reliability, both data and the developed script to process data are available on https://github.com/greendiscbio/neuro_MiningAndModeling/tree/Diagnostic_aid_model on request from computational and clinical researchers [1], where all necessary explanations are provided.

The aim of this work is the development of the computational framework, which is widely customisable and scalable. In this publication, we do not target the accuracy of the clinical assessment provided by the tool and presented in publications like [27], but we discuss around a case of study to show the functionalities of the AI-based tool.

This computational framework has been evaluated using a dataset, which includes cognitive and PET data from 329 patients (171 AD, 72 bvFTD and 87 Healthy controls. As this work is focused on the presentation of the computational framework, we use our own dataset because the data labelling is controlled. Although, this framework has been designed to be able to work with publicly available datasets. This comprehensive dataset is structured using different combinations in order to present a comprehensive and

consistent study. Patients included in this study had a neuroimaging compatible with FDG-PET meeting the current diagnostic criteria [2, 38, 46]. The diagnosis was confirmed after over two years of follow-up. Spouses and volunteers were recruited as Healthy Controls meeting the following criteria: (1) absence of cognitive impairment, according to a MMSE score ≥ 27 and Clinical Dementia Rating of 0 (Morris, 1993); (2) absence of functional impairment measured by Functional Activities Questionnaire scores of 0 [40]. The exclusion criteria were as follows: (1) prior or current history of other neurological diseases (e.g. stroke, brain tumour, seizures); (2) history of psychiatric disease, alcohol or psychotropic drugs abuse; (3) visual, hearing, or any physical problem with a negative impact on test performance.

Regarding data, the Institutional Research Ethics Committee from Hospital Clinico San Carlos approved the research protocol with the 1964 Helsinki declaration and its later amendments. Informed consent was obtained from all individual participants included in the study or their caregivers.

Once the dataset is defined the data preprocessing and feature selection tasks are carried out. Subsequently, AI-based modelling strategies can be launched, and finally, analysing the results obtained through the available metrics in this framework.

3 Results

This section presents the framework design. The code is made available through the *GitHub* and *Pypi* platforms. Figure 1 represents the general scheme of the proposed framework, which is divided into three different parts: (i) Data pre-processing, (ii) Feature engineering, and (iii) AI-based modelling.

3.1 Data pre-processing

Considering the specifications in Section 2, the database has been structured according to the following layout, each brain atlas has two associated tables, one with hypometabolism quantitative data and the other with qualitative data. On the other hand, cognitive evaluations are subdivided into screening and specific tests. Within the specific test there are either raw scores (specific_raw) or scores corrected according to gender, age and years of education (specific_corrected).

Data pre-processing includes all the tasks described below.

Data cleaning This task analyses data and eliminates variables that are neither irrelevant o implicit in the data. Thus, PET date, date of birth, age of disease onset, date of visit, read/write and Mini Mental State Examination (MMSE)

---

[1] Data that support this study are available on request from researchers

were excluded. It also examines the brain data for inconsistencies or incoherencies, e.g. 9 instances with normal brain metabolism which are classified as AD or FTD patients. This information is presented to the user in order to request an action on those instances and/or variables..

Processing of missing values This task is responsible for identifying empty values from the available dataset. It also handles the missing data imputation task. The applied imputation techniques depend on each given dataset and prediction model to be used, so its applicability to another dataset should be analysed. In this framework, missing values imputation was carried out using the non-parametric MissForest imputation technique [50], which is able to identify non-linear and complex relationships between variables. MissForest is an extension of the MICE methods that apply a multivariate and iterative imputation [4], and gives more realistic results than other parametric techniques [50].

Categorization of nominal variables This task is responsible for applying encoding techniques to nominal variables. One Hot Coding is the most frequently used coding scheme, which transforms a single variable with "n" different values into "n" binary variables. Each binary variable represents a single value and the presence is indicated with a 1 and the absence with a 0.

Since the first step of the analysis consisted of a selection of characteristics and each variable in the one hot vector represents a new characteristic, it was not necessary to remove a variable to avoid multi-collinearity problems.

3.2 Feature selection

In high dimensionality problems, identifying the most relevant attributes is a crucial step when modelling data through ML and the problem is an NP problem [13]. Reducing the dimensionality enhances interpretability, a key aspect under the XAI perspective, makes clinical diagnosis easier, improves the performance of classification models, reduces the computational cost and prevents the models from overfitting [48]. This task aims to remove irrelevant and overlapping features from the whole set of features, while retaining the most relevant ones. Hybrid approaches using wrapping techniques, and heuristic and metaheuristic search strategies [32, 59] are very efficient to explore the feature space. Feature selection via evolutionary algorithms [58], as one of the most popular metaheuristic, is selected for the implemented computational tool.

This AI-based tool integrates the feature selection through the **PyWinEA** module, a *Python* package developed on the top of the scikit-learn library that implements the most widely used genetic algorithms. This module is capable of working with data provided by current evaluation and diagnostic techniques. **PyWinEA** package has been endowed with a basic GA and MOEA (NSGAII) to explore the feature space. These techniques and their use along this work are introduced below.

3.2.1 Evolutionary algorithms

Evolutionary algorithms (EA) are population-based techniques inspired by the process of natural selection. They evolve a population of individuals, that represent potential solutions. Individuals will experiment variations to simulate the genetic changes, which guide the evolutionary process.

EAs show a high exploratory capacity, including discontinuous search spaces with a lower tendency for local maxima. This work considered a maximization problem given the interest in improving the models performance. The *PyWinEA* package defines the genotype of the individual as an array of integer values of variable length. Each integer represents an attribute, and the mapping process consists of substituting the integer with the values associated with the attribute. The fitness function is given by the classification model and its classification performance. *PyWinEA* implements two stochastic selection operators: fitness proportional selection and tournament selection, and two survivor selection strategies: elitism and annihilation. Finally, the mutation and recombination operators, random resetting and one-point crossover, are implemented as variation operators.

3.2.2 Multiobjective evolutionary algorithms

Most real problems require more than one metric to evaluate the quality of a potential solution. Frequently, there are several objectives to maximize and usually, they are conflicting objectives. The optimal solutions in multi-objective optimisation deal with the domination concept, which determines the non-dominated front of solutions also called Pareto's front [19, 195–198]

Consequently, when there are two objective functions that are contradictory (e.g. the classification performance and the number of characteristics in the subset), a unique solution may not dominate the rest. In this situation, we are interested in finding the set of non-dominated solutions that are closest to the optimal Pareto's front.

One of the most used MOEAs is the NSGAII [15], which has been implemented in *PyWinEA*. Solutions in the optimal Pareto's front were evaluated by the hypervolume indicator (I_H), which has been applied using the inclusion-exclusion algorithm [57]. I_H is a unitary measure defined in [5] as "the d-dimensional volume of the hole-free orthogonal polytope".

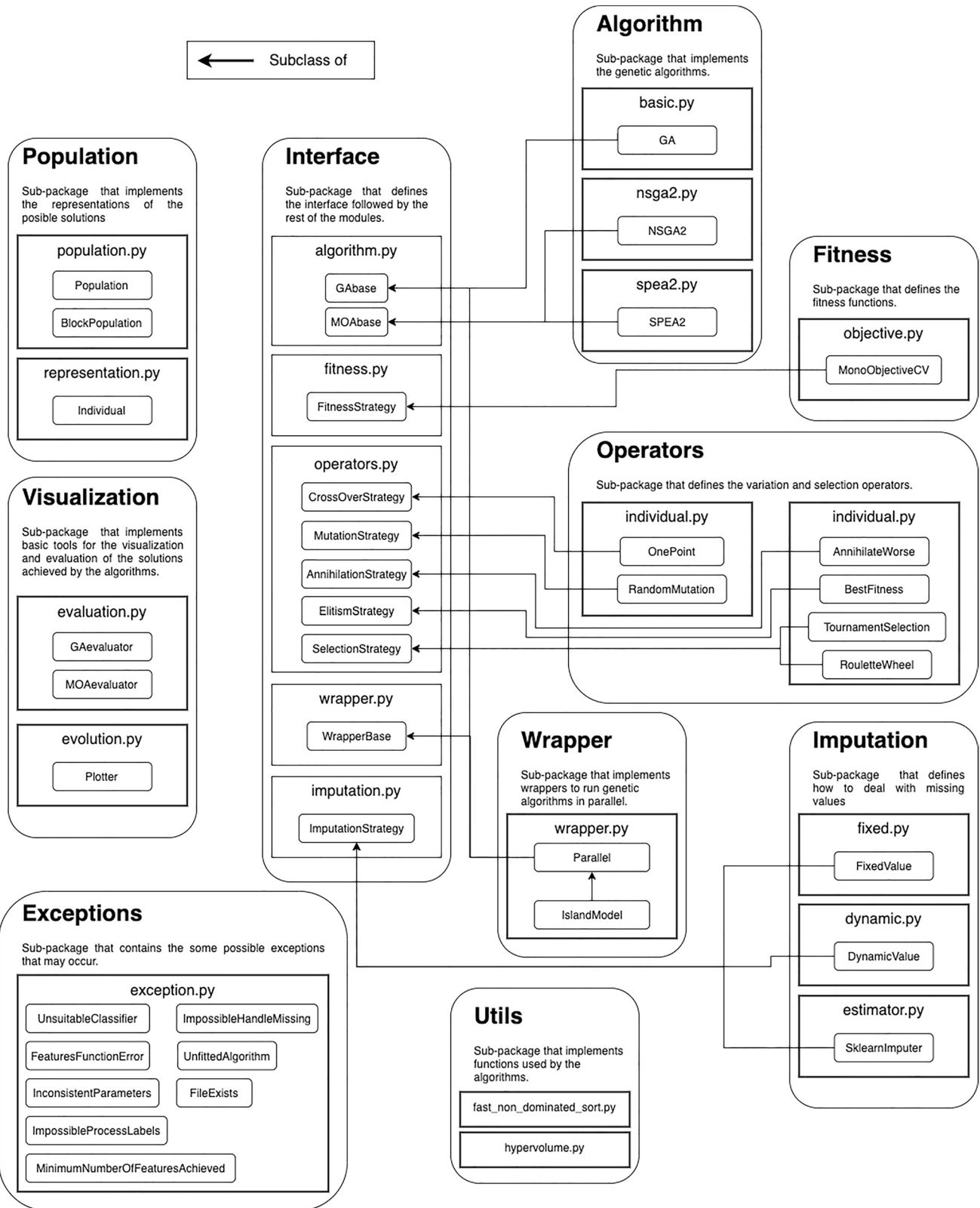A set of supervised classification algorithms has been used to evaluate the quality of the solutions in the feature

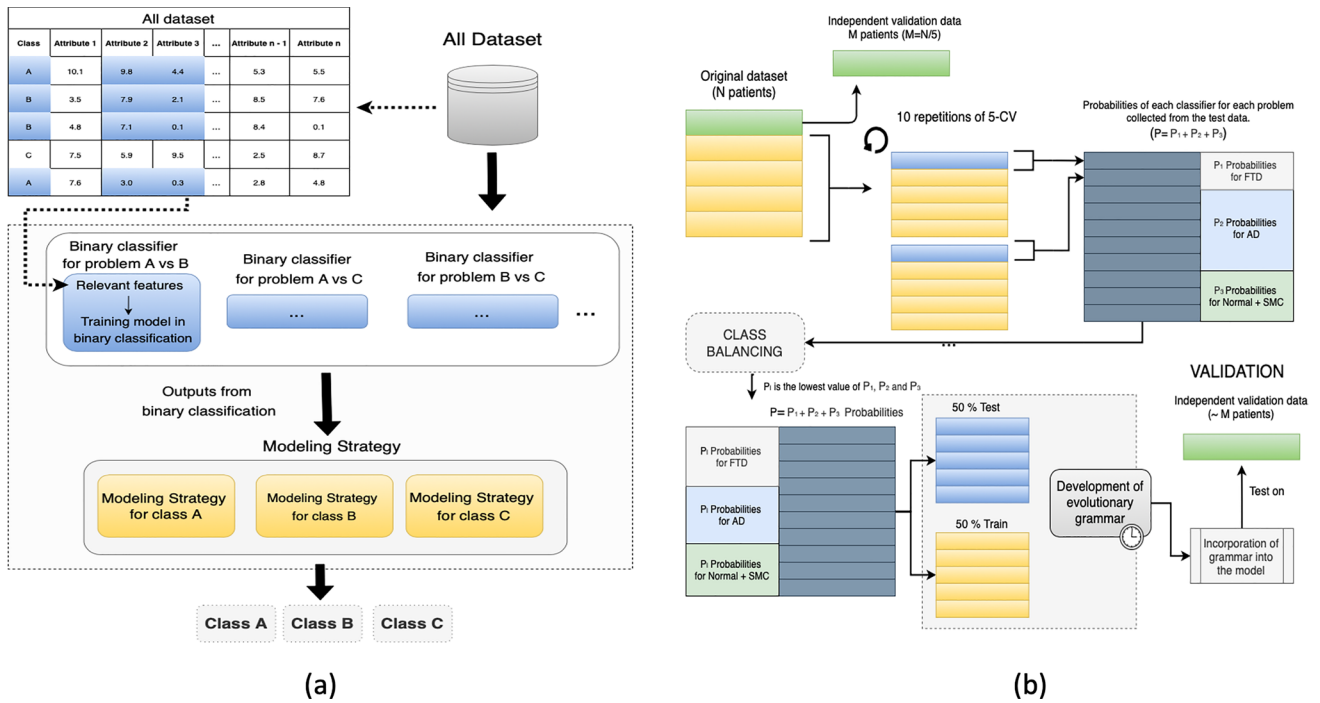**Fig. 2** Structure of the PyWinEA package used for feature selection This package is available through PyPi and GitHub

Fig. 3 Meta-model scheme considering a problem with three classes *A*, *B* and *C*. The modeling strategy takes the output of the binary classifiers of the previous layer and the class assigned to an example will be the one with the highest value

selection process implemented in *PyWinEA* Fig. 2, and to develop (ML)-based solutions.

3.3 ML-based solutions

This section presents the methodology used in developing several learning models to assist clinicians in the diagnosis of AD and FTD.

3.3.1 Machine learning models

This computational tool integrates several classification models to provide clinicians with a widely comparative framework. In this light, different classifiers and their performance can be analyzed using the features selected by the EA algorithm approach. Although any parameter of the classification algorithms can be adjusted, for each classifier we
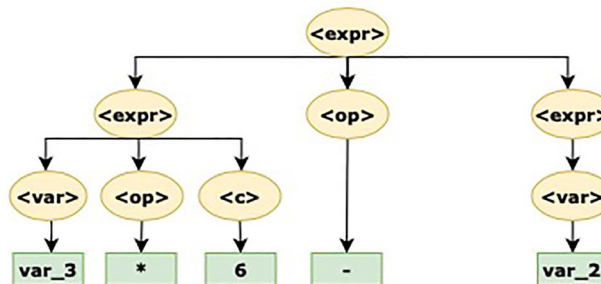

Fig. 4 Genotype to phenotype mapping process following the syntax described in the grammar. The next node to be chosen during the mapping process is determined by the genotype codon module and ends when a terminal node is reached

Fig. 5 Grammar to handle the genotype to phenotype mapping process

```
N = { <op>, <var_1>, <var_2>, <var_3>, <var_4>, <list_vars>, <idx>,
      <func>, <func2>, <exponent>, <sign>, <const>, <cat>, <digit> }

T = { x[0], x[1], x[2], x[3], x[4], log, sqrt2, exp, sin, cos,
      1/, mean, std, min, max, - , +, *, /, 1...4,  1...9 }

S = { <var_0>, <op>, <var_1>, <op>, <var_2>, <op>, <var_3>, <op>, <var_4> }

P = {
    I.     <var_1> ::= x[0] | <var_0> <op> <var> | x[0] | <func>(x[0]) | x[0] |
           power(x[0], <exponent>) | x[0] | <func2>([x[0], <list_vars>])
    II.    <var1> ::= x[1] | <var_1> <op> <var> | x[1] | <func>(x[1]) | x[1] |
           power(x[1], <exponent>) | x[1] | <func2>([x[1], <list_vars>])
    III.   <var_2> ::= x[2] | <var_2> <op> <var> | x[2] | <func>(x[2] | x[2] |
           power(x[2], <exponent>) | x[2] | <func2>([x[2], <list_vars>])
    IV.    <var_3> ::= x[3] | <var_3> <op> <var>| x[3] | <func>(x[3]) | x[3] |
           power(x[3], <exponent>) | x[3] | <func2>([x[3], <list_vars>])
    V.     <var_4> ::= x[4] | <var_4> <op> <var> | x[4] | <func>(x[4]) | x[4] |
           power(x[4], <exponent>) | x[4] | <func2>([x[4], <list_vars>])
    VI.    <op> ::= + | - | * | /
    VII.   <var> ::= <var_0> | <var_1> | <var_2> | <var_3> | <var_4>
    VIII.  <list_vars> ::= x[<idx>] | x[<idx>], <list_vars>
    IX.    <idx> ::=  0 | 1 | 2 | 3| 4
    X.     <func> ::= log | sqrt2 | exp | sin | cos | 1/
    XI.    <func2> ::= mean | std | min | max
    XII.   <exponent> ::= (<sign> <const>) | (<sign> <var>) | <func>(<var>)
    XIII.  <sign> ::= - | +
    XIV.   <const> ::= <sign> <cat>.<cat>
    XV.    <cat> ::= <cat> <digit> |  <digit>
    XVI.   <digit> ::= 0 | 1 | ... | 9
}
```

only mention the most significant ones when addressing this particular problem.

- **Bernoulli naive Bayes**. This model allows to adjust the prior probabilities of each class and the smoothing of the variance.
- **Support Vector Machines**. The RBF (Radial Basis Function) was used as a kernel function and the γ and C parameters were adjusted.
- **K-Nearest Neighbors**. Different number of neighbours and distance metrics were explored.
- **Decision Trees**. Alternative ways of partitioning the nodes (using the best split given by the Gini criterion or by randomly partitioning the nodes), the maximum depth, the minimum number of samples in each split and the minimum number of samples to declare a node as a leaf were the adjusted hyperparameters.

In addition, three ensembles based on decision trees were used. For each one, the number of base estimators and their hyperparameters were tuned:

- **Random Forest**.
- **AdaBoost**. Different learning rates were considered.
- **Gradient Boosting**. The learning rate, the fraction of samples used to train each of the estimators as well as the loss function were adjusted.

Four functionalities were also developed: (1) Graphical evaluation using training and validation; (2) General functionalities such as loading datasets and exception control; (3) Graphical representations among several classification models; (4) Performance evaluation using accuracy, F1-score, precision, recall, sensitivity and specificity metrics, as well as receiver operating characteristic (ROC) curve and classification errors.

Using these functionalities, every classification model provides graphical resources to evaluate the performance of the results, thus the confusion matrix, accuracy, F1-score, precision, recall, learning rate, sensitivity, specificity, the area under Receive Operating Characteristics (ROC) curve and the classification errors are graphically represented.

Fig. 6 Methodology designed for the development and validation of EG as modeling strategy. A class stratification following a CV scheme was implemented. If classes were imbalanced, classes would be balanced to the minority class
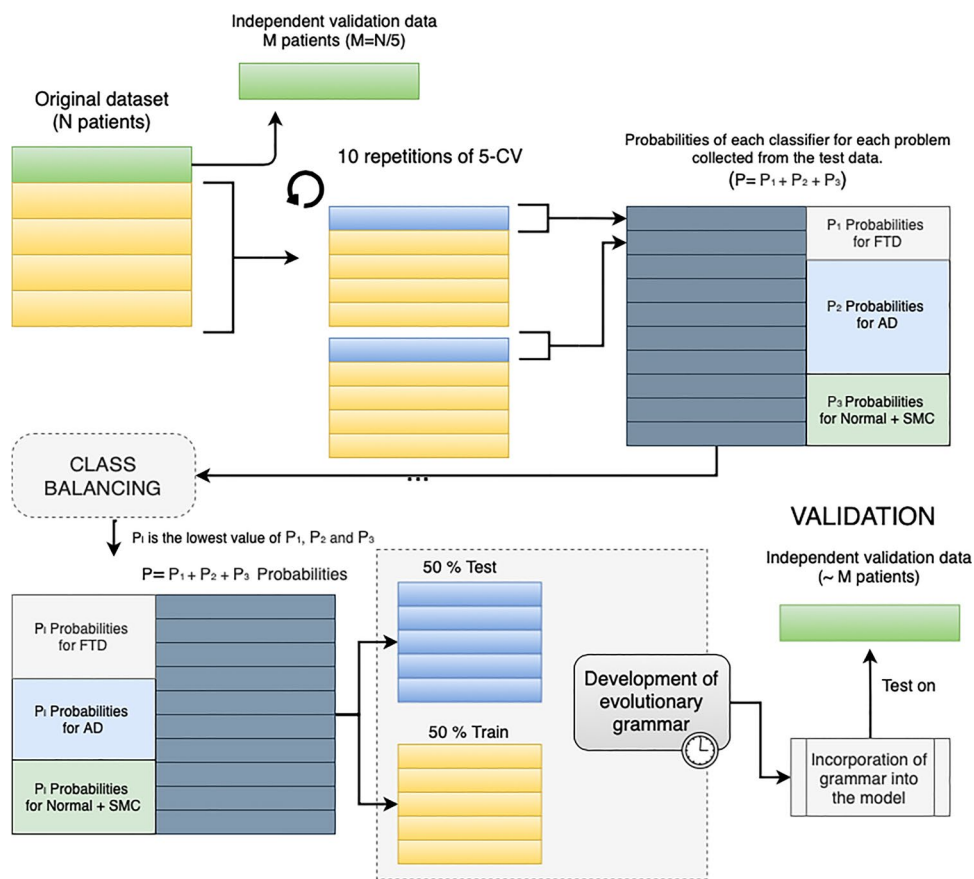
The proposed classifiers cover most of the problems that can be defined with the data processed in the Section 3.1 section. Moreover, a new multiclass classification strategy for cognitive tests is described below.

3.3.2 Meta-model strategy

This work explores a new high quality strategy to improve the classification performance especially designed for cognitive tests when tackling *One vs Rest* problems. It integrates the information provided by each binary classifier into a multiclass single model.

The proposed meta-model is a two-layers design, as presented in Fig. 3, according to a stacking strategy [56]. The first layer is responsible for the binary classification, operating in a different feature space and using characteristics selected during the feature engineering process. This layer uses SVMs as binary classifiers and forwards their results to the second layer that generates a multiclass output. The second layer applies a modeling strategy based on evolutionary grammars or Bayesian networks. In this model, each of the binary classifiers of the first layer operates in a different feature space. Features selected during the feature selection phase were used. Additionally, every binary classifier was trained using different examples, which were driven by the binary problem addressed.

Every target class is associated with one or more binary classifiers. The classification process generates a probability or binary value, which indicates the class that a sample

Table 1 Parameters used for the development of evolutionary grammars using PonyGE2

Parameter	Parameter setting
Algorithm	NSGAII [15]
Population size	300
Elite size	30
Generations	1500
Crossover	subtree[a]
Crossover probability	0.9
Mutation	subtree[a]
Mutation events	1
Selection proportion	0.5
Fitness 1	F1
Fitness 2	Minimizing the number of nodes
Maximum derivation tree initialization depth	10
Maximum derivation tree depth	15
Initialization strategy	PI grow [21]

[a] The crossover strategy is analogous to the one-point operator but by mixing tree structures. The mutation operator is applied only to the population resulting from the crossover

brodmann_qualitative ▼
- 🔑 id INT
- ◇ brodmann_1 FLOAT(10,2)
- ◇ brodmann_2 FLOAT(10,2)
- ◇ brodmann_3 FLOAT(10,2)
- ◇ brodmann_4 FLOAT(10,2)
- ◇ brodmann_5 FLOAT(10,2)
- ◇ brodmann_6 FLOAT(10,2)
- ◇ brodmann_7 FLOAT(10,2)
- ◇ brodmann_8 FLOAT(10,2)
- ◇ brodmann_9 FLOAT(10,2)
- 38 more...
- Indexes ▶

brodmann_quantitative ▼
- 🔑 id INT
- ◇ brodmann_1 FLOAT(10,2)
- ◇ brodmann_2 FLOAT(10,2)
- ◇ brodmann_3 FLOAT(10,2)
- ◇ brodmann_4 FLOAT(10,2)
- ◇ brodmann_5 FLOAT(10,2)
- ◇ brodmann_6 FLOAT(10,2)
- ◇ brodmann_7 FLOAT(10,2)
- ◇ brodmann_8 FLOAT(10,2)
- ◇ brodmann_9 FLOAT(10,2)
- 38 more...
- Indexes ▶

aal_quantitative ▼
- 🔑 id INT
- ◇ pre_l FLOAT(10,2)
- ◇ pre_r FLOAT(10,2)
- ◇ post_l FLOAT(10,2)
- ◇ post_r FLOAT(10,2)
- ◇ ro_l FLOAT(10,2)
- ◇ ro_r FLOAT(10,2)
- ◇ f1_l FLOAT(10,2)
- ◇ f1_r FLOAT(10,2)
- ◇ f2_l FLOAT(10,2)
- 81 more...
- Indexes ▶

aal_qualitative ▼
- 🔑 id INT
- ◇ pre_l FLOAT(10,2)
- ◇ pre_r FLOAT(10,2)
- ◇ post_l FLOAT(10,2)
- ◇ post_r FLOAT(10,2)
- ◇ ro_l FLOAT(10,2)
- ◇ ro_r FLOAT(10,2)
- ◇ f1_l FLOAT(10,2)
- ◇ f1_r FLOAT(10,2)
- ◇ f2_l FLOAT(10,2)
- 81 more...
- Indexes ▶

identifiers ▼
- 🔑 id INT
- ◇ cd_pet FLOAT(10,2)
- Indexes ▶

dates ▼
- 🔑 id INT
- ◇ birth_date DATE
- ◇ pet_date DATE
- ◇ visit_date DATE
- Indexes ▶

diagnostic ▼
- 🔑 id INT
- ◇ diagnostic FLOAT(10,2)
- Indexes ▶

screening ▼
- 🔑 id INT
- ◇ mmse FLOAT(10,2)
- ◇ addensbrook FLOAT(10,2)
- ◇ ace3_total FLOAT(10,2)
- ◇ ace3_attention FLOAT(10,2)
- ◇ ace3_memory FLOAT(10,2)
- ◇ ace3_fluidity FLOAT(10,2)
- ◇ ace3_language FLOAT(10,2)
- ◇ ace3_visospatial FLOAT(10,2)
- Indexes ▶

demographic ▼
- 🔑 id INT
- ◇ sex FLOAT(10,2)
- ◇ current_age FLOAT(10,2)
- ◇ starting_age FLOAT(10,2)
- ◇ education_years FLOAT(10,2)
- ◇ read_and_write FLOAT(10,2)
- Indexes ▶

specific_raw ▼
- 🔑 id INT
- ◇ mst_direct FLOAT(10,2)
- ◇ mst_indirect FLOAT(10,2)
- ◇ cbtt_direct FLOAT(10,2)
- ◇ cbtt_indirect FLOAT(10,2)
- ◇ tmt_a FLOAT(10,2)
- ◇ tmt_b FLOAT(10,2)
- ◇ sdmt FLOAT(10,2)
- ◇ bnt FLOAT(10,2)
- ◇ fcsrt_l1 FLOAT(10,2)
- ◇ fcsrt_lt FLOAT(10,2)
- ◇ fcsrt_total FLOAT(10,2)
- ◇ fcsrt_dif_free FLOAT(10,2)
- ◇ fcsrt_dif_total FLOAT(10,2)
- ◇ st_1 FLOAT(10,2)
- ◇ st_2 FLOAT(10,2)
- ◇ st_3 FLOAT(10,2)
- ◇ tlt_total_correct FLOAT(10,2)
- ◇ tlt_total_movements FLOAT(10,2)
- ◇ tlt_start FLOAT(10,2)
- ◇ tlt_execution FLOAT(10,2)
- ◇ tlt_resolution FLOAT(10,2)
- ◇ vospt_decision FLOAT(10,2)
- ◇ vospt_silhouette FLOAT(10,2)
- ◇ vospt_discrimination FLOAT(10,2)
- ◇ vospt_location FLOAT(10,2)
- ◇ jlot FLOAT(10,2)
- ◇ ft_formal FLOAT(10,2)
- ◇ ft_animals FLOAT(10,2)
- ◇ rocf_copy FLOAT(10,2)
- ◇ rocf_3min FLOAT(10,2)
- ◇ rocf_30min FLOAT(10,2)
- ◇ rocf_time FLOAT(10,2)
- ◇ rocf_recognition FLOAT(10,2)
- ◇ rocf_type_copy FLOAT(10,2)
- ◇ rocf_type_3min FLOAT(10,2)
- ◇ rocf_type_30min FLOAT(10,2)
- ◇ pctb FLOAT(10,2)
- Indexes ▶

specific_corrected ▼
- 🔑 id INT
- ◇ mst_direct FLOAT(10,2)
- ◇ mst_indirect FLOAT(10,2)
- ◇ cbtt_direct FLOAT(10,2)
- ◇ cbtt_indirect FLOAT(10,2)
- ◇ tmt_a FLOAT(10,2)
- ◇ tmt_b FLOAT(10,2)
- ◇ sdmt FLOAT(10,2)
- ◇ bnt FLOAT(10,2)
- ◇ fcsrt_l1 FLOAT(10,2)
- ◇ fcsrt_lt FLOAT(10,2)
- ◇ fcsrt_total FLOAT(10,2)
- ◇ fcsrt_dif_free FLOAT(10,2)
- ◇ fcsrt_dif_total FLOAT(10,2)
- ◇ st_1 FLOAT(10,2)
- ◇ st_2 FLOAT(10,2)
- ◇ st_3 FLOAT(10,2)
- ◇ tlt_total_correct FLOAT(10,2)
- ◇ tlt_total_movements FLOAT(10,2)
- ◇ tlt_start FLOAT(10,2)
- ◇ tlt_execution FLOAT(10,2)
- ◇ tlt_resolution FLOAT(10,2)
- ◇ vospt_decision FLOAT(10,2)
- ◇ vospt_silhouette FLOAT(10,2)
- ◇ vospt_discrimination FLOAT(10,2)
- ◇ vospt_location FLOAT(10,2)
- ◇ jlot FLOAT(10,2)
- ◇ ft_formal FLOAT(10,2)
- ◇ ft_animals FLOAT(10,2)
- ◇ rocf_copy FLOAT(10,2)
- ◇ rocf_3min FLOAT(10,2)
- ◇ rocf_30min FLOAT(10,2)
- ◇ rocf_time FLOAT(10,2)
- ◇ rocf_recognition FLOAT(10,2)
- Indexes ▶

◄**Fig. 7** The structure of the database designed. Tables brodmann_qualitative/quantitative and aal_qualitative/quantitative, corresponding to the brain metabolism data have been shortened. Complete data are available on request on GitHub

belongs to. The modeling strategy associates the results of the binary classifiers to a single real value. The highest value will identify the final class.

Figure 3 considers a problem with three classes *A*, *B* and *C* and three binary classifiers C_{AvsB}, C_{AvsC} and C_{BvsC}, which use different characteristics to perform the classification. Given a training dataset *T*, the first step consists on the generation of three datasets T_1, T_2 and T_3. The dataset T_1 associated with C_{AvsB} is composed of the characteristics selected for the *A* vs *B* problem and the examples labelled with classes *A* and *B* excluding the examples belonging to *C*. The same is applied to datasets T_2 and T_3.

During the prediction phase, we will have a modelling strategy associated to each class. The modelling strategy associated to class *A* will receive the outputs of classifiers C_{AvsC} and C_{AvsB}, the same for the rest of the classes and their associated classifiers. The class selected will be decided upon the modelling strategy that provides the highest value.

Evolutionary grammars as a modelling strategy Evolutionary grammars (EG) are part of EAs and an approach to genetic programming. Solutions are generated using a grammar representation. EG has obtained promising results in many domains such as the prediction of migraine crisis [42] or glucose levels [12, 30, 55].

Representing the genotype with an array of integer or binary values, the genotype-to-phenotype decoding uses a Backus Naur Form (BNF) grammar [47]. Figure 4 describes an example of the mapping process. A grammar is represented by the tuple {*N*,*T*,*P*,*S*} where *N* and *T* are the non-terminal and terminal symbols, respectively; *P* are the production rules applied to generate *T* from *N*, and *S* is the initial expression. The result is a tree structure where S represents the root, N the intermediate nodes, P the potential paths and T the leaves.

Figure 5 shows the grammar used to define the genotype-to-phenotype mapping process, where the gender and age variables are not included to avoid bias. The variable *x* refers to the set of predictions made by the binary classifiers of the previous layer, therefore the index indicates the position of the output of the algorithm associated with a given binary problem.

Figure 6 shows the methodology followed by the proposed meta-model using EG as modelling strategy. The steps are described below:

1. The dataset was divided into 5 disjunct datasets with class stratification following a cross-validation (CV) scheme.
2. One of the datasets is reserved independently for the validation process. With the remaining four, the binary classification phase is launched for 10 iterations with a 5-CV scheme. The predictions of the binary classification models generate a new dataset.
3. If classes were unbalanced, at this point they would be balanced to the minority class. This is carried out by randomly removing predictions from the majority classes until all classes are balanced.
4. The grammar development uses 50% of the dataset samples for training and 50% for testing. This process can be defined as a new supervised classification problem
5. This grammar is integrated into the model as a modelling strategy. The validation of the meta-model is performed with the independent dataset from the step 1.
6. The steps from 2 to 5 are repeated for each of the 5 separate folds in step 1.

The described procedure allows to make an approximation of the generalization capacity of the meta-model that incorporates EG in the second layer. Based on the approach of [44] and given its influence on AD [3], the methodology described was repeated after introducing the gender and age variables into the prediction dataset and including them in the grammars. Thus, the production rule ***VII*** was modified to include the gender (x[5]) and age (x[6]) variables:

VII. <var> ::= <var_0> | <var_1> | <var_2> | <var_3> | <var_4> | x[5] | x[6]

The grammar was implemented using the Python package PonyGE2 [22]. Table 1 shows the default selected parameters, although other parameter values can be applied.

Bayesian networks as a modelling strategy Bayesian networks represent a sub-type of probabilistic graphical models. This type of model uses directed acyclic graphs (DAG) to represent the probabilistic relationship between variables. Nodes correspond to variables and an arc between two nodes shows the dependency relationship. In this type of models, every node is associated to a local probability distribution, which is usually specified by a conditional probability table (CPT), and depends on its parents [33, 42–92]. Each node receives an input and gives the probability distribution of the variable associated to the node, as an output [2].

---

[2] Supplementary Material is provided for this multiclass meta-model

Table 2 Selected parameters for the NSGAII algorithm used to carry out the feature selection
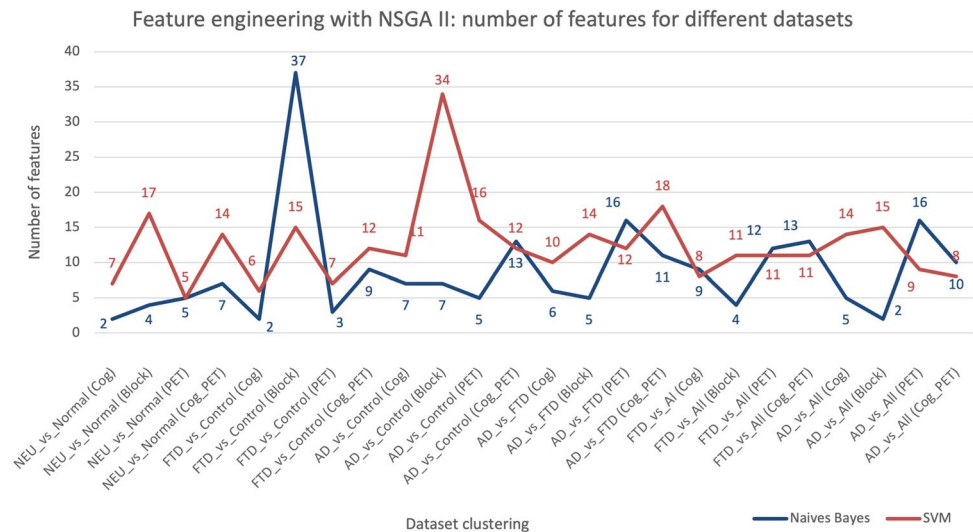
Algorithm parameter	Parameter setting
Mutation strategy	Random Resetting
Selection strategy	Tournament selection[a]
Fitness 1	Accuracy or F1[b]
Fitness 2	Number of features[c]
Number of different initializations	2

[a] k = 2, winners = 1 without replacement

[b] 5 repetitions of 5-CV with class stratification

[c] Defined by equation: $1 - \frac{\text{Length (Individual)}}{\text{Num. Features}}$

Fig. 8 Comparative results of the Feature engineering using NSGAII, Naive Bayes and SVM algorithms as the fitness function. X axis represents the experiments addressed, and Y axis is the number of features obtained. Block means Demographic data + Cognitive test data without separating the scores associated with the same test

Feature engineering with NSGA II: number of features for different datasets

Bayesian network, given the evidence (binary classifier outputs), yields the maximum probability.

4 Discussion in a case of study

This section presents some outcomes that can be achieved by the proposed framework in a particular case study of neurological diseases: clinical diagnosis of AD and FTD. A description of data in this study is presented in Section 2. Although, this study is not focused on the clinical analysis of AD and FTD by means of the proposed framework, we

In this computational tool, the meta-model based on Bayesian networks was implemented on the top of Pomegranate library [49]. Each node in the network corresponds to a binary classifier associated with a given problem. Thus, the Bayesian network allows to model the joint probability distribution of the output of the binary classifiers by assigning a probability to each possible combination of outputs. The two steps required to build a Bayesian network include learning the structure and determining the probability distribution associated with each node based on the data. The structure was determined using a score-base approach, applying dynamic programming and the A* algorithm in order to maximize the probability of the data given the model by means of maximum likelihood estimation.

The dataset generated by the grammars during the step 1 was used for the learning network. Predictions were binarized by rounding up to the nearest integer. A Bayesian network was developed to model the joint probability for each of the classes in such a way that, the label assigned to a new example corresponds to the class whose associated

present a case of study using PET data in order to show the functionalities of the tool. Particularly, data preprocessing phase, feature engineering using NSGAII, classification using different ML algorithms, multiclass meta-model with EG and Bayesian networks, and some of the graphical resources to outline the results. We expect that, with this case of the study, the reader will understand the capabilities of the proposed computing framework and will be able to value the potential of the tool in its clinical practice.

4.1 Data preprocessing

Regarding the data pre-processing described in Section 3.1, data were structured in a relational MySQL database shown in Fig. 7, which can be extendable as needed.

Once the database was ready, data were analysed within the cleaning process and removed irrelevant data. Then, the analysis of the missing values was carried out, which represented 11.28% in the database. After its identification, the Missforest imputation technique was applied with 100

Table 3 *AD − FTD* vs HC: Features selected using NSGA II with SVM as fitness function for Cognitive, Block, PET and *PET* +Cognitive

Cognitive	Block	PET	*PET* +Cognitive
ace3_fluidity	education_years	o1_l	f1m_l
rocf_type_3min_4	cbtt_direct	brodmann_47	brodmann_35
mst_direct	cbtt_indirect	f1m_l	pcl_r
tmt_a	tmt_a	brodmann_35	sma_l
fcsrt_lt	tmt_b	brodmann_37	put_r
rocf_30min	fcsrt_l1		cau_l
education_years	fcsrt_lt		brodmann_19
	fcsrt_total		fcsrt_dif_free
	fcsrt_dif_free		ace3_total
	fcsrt_dif_total		fcsrt_dif_total
	addensbrook		sdmt
	ace3_total		rocf_type_copy_4
	ace3_attention		rocf_type_3min_7
	ace3_memory		rocf_type_copy_3
	ace3_fluidity		
	ace3_language		
	ace3_visospatial		

as the maximum number of interactions and the following parameters settings: Mean for the initial imputation; $1e − 03$ as early stopping; 50 as number of trees (default parameters for decision trees); Mean squared error as the evaluation criterion of each partition, and random for splitting each node.

The last step was dealing with nominal variables following the methodology presented in Section 3.1. The imputation generates real values, which will be rounded to the nearest integer. Next, the one hot coding scheme is applied and as many variables as different values were added.

4.2 Features engineering

The aforementioned use case was addressed by bi-objective MOEA approach, previously described in Section 3, and a customization of hyperparameters as shown in Table 2, where the two objective fitness function are also described. 10 iterations of 5-CV were run and the performance of the best subset obtained for each classifier was evaluated. The algorithm were run for 10 iterations of 5-CV and the performance of the best subset obtained for each classifier was evaluated.

The NSGAII MOEAs obtain several solutions as part of the Pareto Front. The set of features selected for each potential solution can be visualized by the physicians to validate the clinical impact. Figure 8 shows results for several datasets: Demographic, Cognitive Test, and Brain Metabolism Data. Table 3 shows the features selected with SVM as fitness function for AD and FTD vs HC. In this example, Bayesian classifiers obtain the best results, with an average reduction of features of 91.52% compared to 87.97% for SVMs classifiers. Considering that the reduction percentage reached for both classifiers is really high, it is necessary to evaluate the performance each individual with Bayesian and SVMs classifiers as fitness function.

The solutions provided by the feature engineering approach are fed to the ML-based phase: classifiers and meta-model using EG as in Fig. 1, described in Section 3. For each problem, only one of the solutions provided in the feature selection phase has been selected for testing. Accuracy and F1-score as more qualified metrics have been selected for the analysis. We remind the reader that this work does not focus on the clinical analysis but on the possibilities opened by the developed tool. Hence, selected problems from the case of study will be presented in order to evaluate such capabilities of the computational tool.

As a result of the evaluation tests, the SVMs classifiers performed slightly better than the Bayesian classifiers as shown in Fig. 9. Table 4 presents the average values of the metrics used to assess the solutions from the Pareto front obtained with NSGA II, and applied in this case of study. High values of sensitivity and specificity indicate the reliability in predicting positive and negative cases, respectively. The significance of these results is evaluated using the $p − value$ in Table 5. Very small $p − values$ confirm the reliability of the study. According to the results, our ML-based tool is able to clearly differentiate between individuals with AD, FTD and healthy controls, especially when PET data are provided. A slightly lower performance is obtained working with cognitive dataset. Although cognitive test performance is closely associated with the brain metabolism of some regions, not all brain regions are covered during the neuropsychological examinations [16, 31, 37]. In addition, other factors such as cognitive reserve may limit the

Fig. 9 Classification performance achieved by one of the best feature subsets given by the NSGAII for each of the algorithms used to evaluate the fitness, when individuals with AD, FTD and healthy controls were evaluated. *Cognitive C.* denotes groupings of scores from the same cognitive test; *Cognitive. I.* considers each of the scores independently
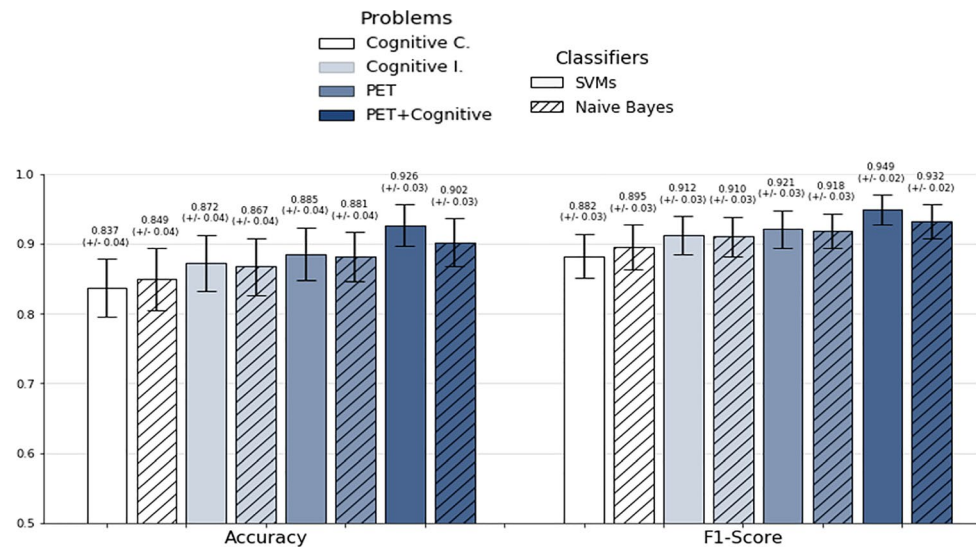

Table 4 Pareto front assessment resulting from NSGA II

Dataset	Accuracy	Precision	Sensitivity	Specificity	F1−score
Naive Bayes					
Cognitive C.	0.849	0.914	0.879	0.766	0.895
Cognitive I.	0.867	0.908	0.913	0.741	0.910
PET	0.881	0.919	0.918	0.782	0.918
PET + Cognitive	0.902	0.954	0.912	0.875	0.932
Support Vector Machine					
Cognitive C.	0.837	0.941	0.832	0.852	0.882
Cognitive I.	0.872	0.927	0.898	0.800	0.912
PET	0.885	0.919	0.924	0.782	0.921
PET + Cognitive	0.926	0.966	0.933	0.906	0.949

diagnostic capacity of neuropsychological examination in some cases [51].

Moreover, this tool provides information about the evolution of the feature engineering process by means of a graphical representation of the evolution of the convergence of the MOEAs. For instance and related to the case of use, Fig. 10 represents the MOEA convergence for the PET datatests addressed using Naive Bayes and SVMs.

In addition, this tool also supplies different visual support representations to evaluate the performance of the classifiers. It implements the receiver operating characteristic (ROC), the confusion matrix, and a comparative graphical representation of the variation in classification performance among the different classifiers with respect to the best result obtained and with the best feature subset

during the featured engineering phase (Fig. 9). Fig. 11 shows the variation performance for the case of use, where the hyperparameters were adjusted using a grid search strategy. Particularly, regularization parameters λ and C for SVM, the loss function (binomial deviance or exponential), the percentage of examples used to train the base models of the ensemble [3] and the number of characteristics [4]. SVM and Gradient Boosting obtained the best performance with $F1 - score = 0.925$, although the rest of algorithms also reached high values. Table 6 shows the p-values computed from the performance metrics along all iterations. It can be seen that the obtained values are much smaller than 0.05 and we can state that the results are significant.

One of the most important challenges for the clinical experts is the interpretability of the models. In this light, decision tree models have been developed to provide this capability to clinicians. This kind of algorithms provide a clear and simple set of rules that allow to distinguish between different clinical conditions. Figure 12 represents the decision tree for the case of use that we are presenting to give insights

[3] This parameter generates a behavior analogous to bagging helping to reduce the variance of the model

[4] This refers to the attributes used to train each of the base classifiers. This parameter generates a behavior similar to randomization. The parameters taken were $\sqrt{num\_features}$ or $\log_2(num\_features)$

Table 5 P−value for metrics applied to assess results from NSGA II

Dataset	P−value				
	Accuracy	Precision	Sensitivity	Specificity	F1−score
Naive Bayes					
Cognitive Ċ	3.02E-25	1.83E-16	2.76E-21	1.50E-15	9.14E-25
Cognitive I.	1.14E-27	2.27E-24	2.81E-22	3.82E-23	9.54E-28
PET	2.19E-28	1.25E-27	1.72E-19	1.91E-26	9.72E-28
PET + Cognitive	7.10E-29	3.72E-22	3.52E-23	8.59E-21	3.88E-28
Support Vector Machine					
Cognitive C.	3.19E-28	1.37E-13	5.43E-27	8.74E-13	3.84E-27
Cognitive I.	3.41E-27	1.42E-18	2.71E-25	3.54E-17	1.65E-27
PET	2.17E-26	3.80E-27	6.16E-19	1.91E-26	7.95E-26
PET + Cognitive	1.67E-22	6.19E-14	4.91E-19	1.12E-13	2.63E-22
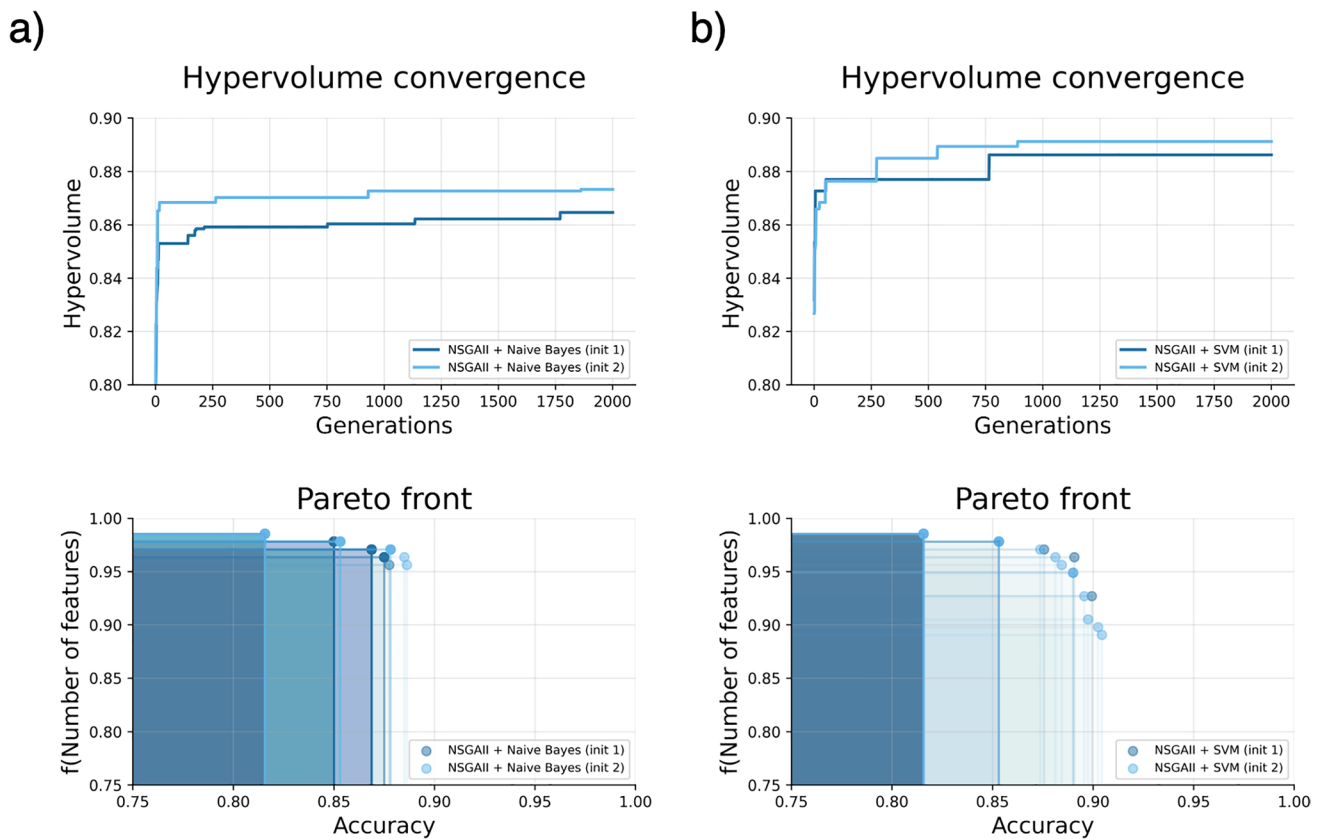
Fig. 10 Convergence of the NSGAII for the Neurodegenerative Disorders (NEU) vs Healthy Controls (HC) diagnosis including PET data using (**a**) Naive Bayes classifier or (**b**) SVMs. NEU represents AD or bvFTD disorders. The pareto front subfigure is defined by equation: $1 - \frac{\text{Length (Individual)}}{\text{Num. Features}}$. The results of two different initialisations are shown

about the functionalities of this framework. This result was validated by expert neurologists who agreed on the clinical significance. According to the expert neurologists, in the decision tree, several key areas in the pathophysiology of AD and/or FTD are included. Specifically, regions in the frontal lobe (frontal superior medial gyrus and inferior frontal gyrus/ Brodmann area 47), the temporal cortex (Brodmann area 37) and occipital lobe. According to the tree, the hypometabolism

of any of these areas suggests the presence of a neurodegenerative disorder, while a normal metabolism in all areas is required to be classified as control.

4.3 Meta-models

This meta-model was designed to work with independent cognitive tests scores. The result is a multiclass

Table 6 *P*-value for metrics applied in the case study

P-value					
Classifier	Accuracy	Precision	Sensitivity	Specificity	F1-Score
Naive Bayes	1.95E-25	1.08E-24	4.05E-17	8.25E-23	6.42E-25
SVM	3.20E-25	2.35E-24	1.96E-17	1.30E-22	3.47E-25
KNN	2.16E-22	1.32E-23	5.38E-16	1.30E-22	4.04E-22
Decision Trees	4.59E-25	3.54E-24	7.41E-19	1.30E-22	1.02E-24
Random Forest	9.36E-25	1.37E-24	2.09E-15	1.30E-22	4.11E-24
Gradient Boosting	3.20E-25	2.35E-24	1.96E-17	1.30E-22	3.47E-25

Fig. 11 Variation in classification performance for the NEU (AD or bvFTD) vs HC diagnosis and PET data. The reference values $Acc = 0.885$, $Pre = 0.919$, $Rec = 0.924$, $F1 = 0.921$ correspond to the highest scores in Fig. 10
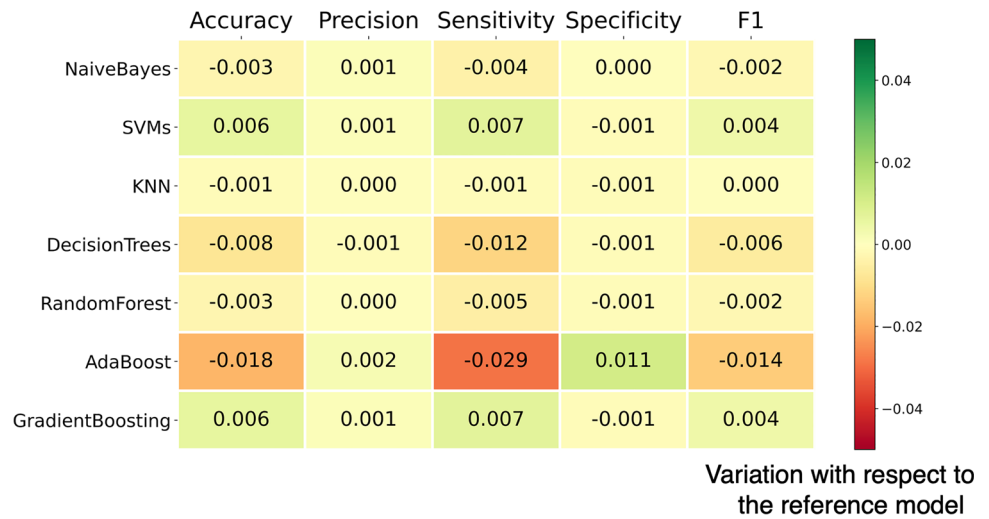
Fig. 12 Decision trees corresponding to the classification problem NEU (AD or bvFTD) vs HC is presented as an example of a more interpretable graph. (Performance: Acc $= 0.885 \pm 0.04$; Pre $= 0.919 \pm 0.03$; Rec $= 0.924 \pm 0.04$; F1 $= 0.921 \pm 0.03$). Squares and ellipses represent nodes and leaves, respectively. The color blue denotes that most instances belong to the class indicated on the leaf but at least 1/4 correspond to the opposite class
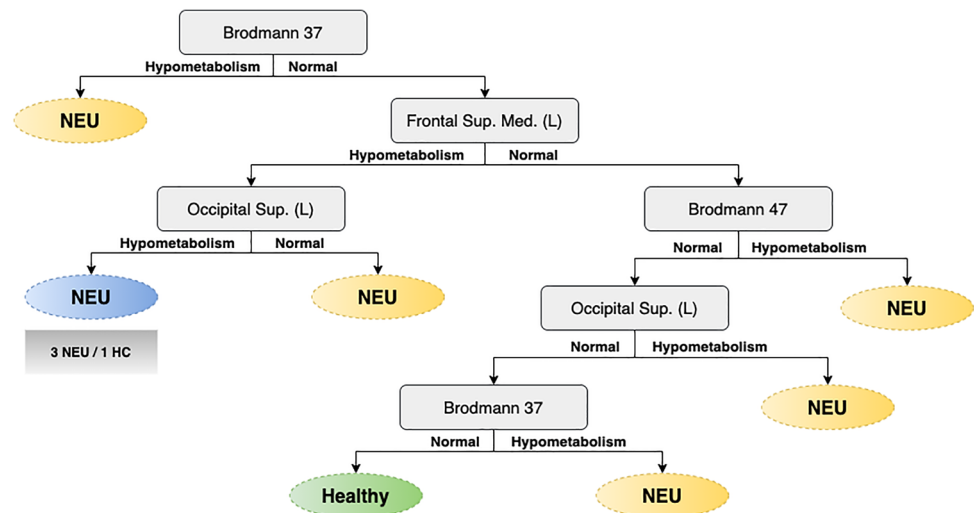

classification model, which integrates the output of the binary classifiers using EG or Bayesian classifiers. In order to validate this module, we show the results obtained using the one vs rest classification models with the AD condition. Figure 13 shows the results obtained using accuracy and F1-score as metrics to evaluate the performance. It is clearly observed that the strategy with EG improved the classification results compared to the best results obtained

using the binary classifiers independently and Bayesian networks. Even after including gender and age variables, which produce a loss of performance, EG overcomes the performance of the previous ones.

Results from this model with EG have demonstrated a great potential to improve the classification accuracy with limited datasets, as cognitive assessments.

Fig. 13 Performance obtained in multi-class classification integrating the output binary classifiers into a multiclass output. Ten repetitions of 5 CV were applied for the validation process of the reference model and the modeling strategy with Bayesian networks (described in Section 3.3.2)
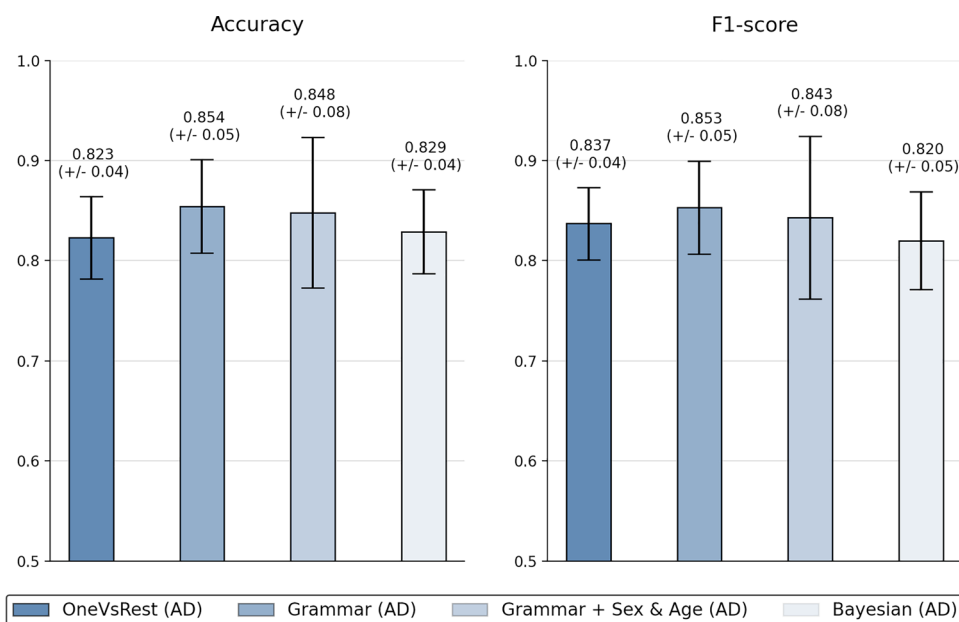
5 Conclusions

This paper has presented the design and implementation of a machine learning–based framework for the automatic diagnosis, especially, of neurodegenerative diseases. Neuropsychological and neuroimaging assessments provide large, heterogeneous datasets, with high possibilities for knowledge mining and the development of diagnostic tools. Our tool is proposed under the XAI perspective to support the clinicians in the diagnosis, as it provides all the steps required to analyse these datasets, from the data preprocessing, feature selection through an evolutionary approach, and modeling of the mentioned diseases.

As a case of study, we have evaluated the performance of our approach in the diagnosis of two widespread neurodegenerative diseases, AD and FTD. It was clearly observed how the proposed framework allows a smooth processing of the cognitive and image assessments, with a high reduction in the number of features needed for the diagnosis, and a high accuracy in the classification. A strong effort has been put on the interpretability of the results, showing how a data-centric point of view helps to understand AD and FTD disorders.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11517-022-02630-z) contains supplementary material, which is available to authorized users.

Data availability Al data are available in a systematic database created by the Department of Neurology of the San Carlos Hospital, in Madrid, and accessible to clinicians and researchers participating in the project. These data are not publicly available due to data privacy laws.

Code availability Code is available in PyPi and GitHub

Declarations

Conflict of interest The authors declare no competing interests.

Ethics approval and consent to participate The Institutional Research Ethics Committee from Hospital Clinico San Carlos approved the research protocol with the 1964 Helsinki declaration and its later amendments. Written informed consent was obtained from all individual participants included in the study or their caregivers.

Consent for publication Not applicable

References

1. Ahmed Z, Mohamed K, Zeeshan S, Dong X (2020) Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database 2020. https://doi.org/10.1093/database/baaa010

2. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC, Snyder PJ, Carrillo MC, Thies B, Phelps CH (2011) The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's Dement 7:270–279

3. Association A (2019) 2019 Alzheimer's disease facts and figures. Alzheimer's & Dementia 15:321–387

4. Azur MJ, Stuart EA, Frangakis C, Leaf PJ (2011) Multiple imputation by chained equations: what is it and how does it work?. Int J Methods Psychiatr Res 20:40–49

5. Beume N, Fonseca CM, Lopez-Ibanez M, Paquete L, Vahrenhold J (2009) On the complexity of computing the hypervolume indicator. IEEE Trans Evol Comput 13:1075–1082

6. Bitam S, Mellouk A (2006) Brodmann's localisation in the cerebral cortex. Springer, Berlin, p 298

7. Brown KJ, Bohnen NI, Wong KK, Minoshima S, Frey KA (2014) Brain pet in suspected dementia: patterns of altered fdg metabolism. Radiographics 34:684–701

8. Brzezicki MA, Kobetíc MD, Neumann S, Pennington C (2019) Diagnostic accuracy of frontotemporal dementia. an artificial intelligence-powered study of symptoms, imaging and clinical judgement. Adv Med Sci 64:292–302. https://doi.org/10.1016/j.advms.2019.03.002

9. Cabitza F, Gensini GF (2017) Unintended consequences of machine learning in medicine. JAMA 318:517518. https://doi.org/10.1001/jama.2017.7797

10. Casanova R, Wagner B, Whitlow CT, Williamson JD, Shumaker SA, Maldjian JA, Espeland MA (2011) High dimensional classification of structural MRI Alzheimer's disease data based on large scale regularization. Front Neuroinformatics 5:22

11. Castro AP, Fernandez-Blanco E, Pazos A, Munteanu CR (2020) Automatic assessment of Alzheimer's disease diagnosis based on deep learning techniques. Comput Biol Med, 103764

12. Contreras I, Oviedo S, Vettoretti M, Visentin R, Veh J (2017) Personalized blood glucose prediction: a hybrid approach using grammatical evolution and physiological models. PLOS ONE 12:1–16. https://doi.org/10.1371/journal.pone.0187754

13. D. S, R. S (1994) Np-completeness of searches for smallest possible feature sets. In: AAAI Symposium on Intelligent Relevance, AAAI Press, pp 37–39

14. De A, Chowdhury AS (2020) Dti based Alzheimer's disease classification with rank modulated fusion of cnns and random forest. Expert Syst Appl 114338. https://doi.org/10.1016/j.eswa.2020.114338

15. Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: Nsga-ii. IEEE Trans Evol Comput 6:182–197

16. Delgado-Álvarez A, Cabrera-Martn MN, Pytel V, Delgado-Alonso C, Matías-Guiu J, Matias-Guiu JA (2021) Design and verbal fluency in Alzheimer's disease and frontotemporal dementia: clinical

and metabolic correlates. J Int Neuropsychol Soc. https://doi.org/10.1017/S1355617721001144, 1–16

17. Ding Y, Sohn JH, Kawczynski MG, Trivedi H, Harnish R, Jenkins NW, Lituiev D, Copeland TP, Aboian MS, Mari Aparici C, Behr SC, Flavell RR, Huang S-Y, Zalocusky KA, Nardo L, Seo Y, Hawkins RA, Hernandez Pampaloni M, Hadley D, Franc BL (2019) A deep learning model to predict a diagnosis of Alzheimer disease by using 18f-FDG PET of the brain. Radiology 290:456–464. https://doi.org/10.1148/radiol.2018180958

18. Dror IE, Kukucka J, Kassin SM, Zapf PA (2018) When expert decision making goes wrong: consensus, bias, the role of experts, and accuracy. J Appl Res Memory Cognit 7:162–163. https://doi.org/10.1016/j.jarmac.2018.01.007

19. Eiben AE, Smith JE (2015) Introduction to evolutionary computing. 2 ed., Springer. https://doi.org/10.1007/978-3-662-44874-8

20. Erkkinen MG, Kim M-O, Geschwind MD (2018) Clinical neurology and epidemiology of the major neurodegenerative diseases. Cold Spring Harbor Perspectives in Biology 10:a033118

21. Fagan D, Fenton M, O'Neill M (2016) Exploring position independent initialisation in grammatical evolution. In: 2016 IEEE Congress on Evolutionary Computation (CEC), IEEE. pp. 5060-5067

22. Fenton M, McDermott J, Fagan D, Forstenlechner S, Hemberg E, O'Neill M (2017) PonyGE2: Grammatical evolution in Python. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion, ACM, Berlin, Germany. pp. 1194-1201

23. Fernández-Matarrubia M, Matías-Guiu JA, Moreno-Ramos T, Matías-Guiu J (2014) Demencia frontotemporal variante conductual: aproximación clínica y terapéutica. Neurología 29:464–472

24. Fernández-Matarrubia M, Matías-Guiu JA, Cabrera-Martín MN, Moreno-Ramos T, Valles-Salgado M, Carreras JL, Matías-Guiu J (2017) Episodic memory dysfunction in behavioral variant frontotemporal dementia: A clinical and fdg-pet study. J Alzheimer's Dis: 1251 1264. https://doi.org/10.3233/JAD-160874

25. Fisher CK, Smith AM, Walsh JR, Simon EA (2019) Machine learning for comprehensive forecasting of Alzheimer's disease progression. Scient Reports 9:13622. https://doi.org/10.1038/s41598-019-49656-2

26. Foster NL, Heidebrink JL, Clark CM, Jagust WJ, Arnold SE, Barbas NR, DeCarli CS, Scott Turner R, Koeppe RA, Higdon R et al (2007) Fdg-pet improves accuracy in distinguishing frontotemporal dementia and Alzheimer's disease. Brain 130:2616–2635

27. Garcia-Gutierrez F, Delgado-Alvarez A, Delgado-Alonso C, Díaz-Álvarez J, Pytel V, Valles-Salgado M, Gil MJ, Hernández-Lorenzo L, Matías-Guiu J, Ayala JL et al (2022) Diagnosis of Alzheimer's disease and behavioural variant frontotemporal dementia with machine learning-aided neuropsychological assessment using feature engineering and genetic algorithms. Int J Geriatr Psychiatry:37

28. Gunning D (2017) Explainable artificial intelligence (xai). http://www.darpa.mil/program/explainable-artificial-intelligence. Accessed 10/11/2021

29. Harper L, Fumagalli GG, Barkhof F, Scheltens P, OBrien JT, Bouwman F, Burton EJ, Rohrer JD, Fox NC, Ridgway GR, Schott JM (2016) MRI visual rating scales in the diagnosis of dementia: evaluation in 184 post-mortem confirmed cases. Brain 139(4):1211–1225. https://doi.org/10.1093/brain/aww005

30. Hidalgo JI, Colmenar JM, Kronberger G, Winkler SM, Garnica O, Lanchares J (2017) Data based prediction of blood glucose concentrations using evolutionary methods. J Med Syst 41:142

31. JA M-G, MN C-M, Valles-Salgado M EA (2017) Neural basis of cognitive assessment in Alzheimer disease, amnestic mild cognitive impairment, and subjective memory complaintsia, amyotrophic lateral sclerosis, and Alzheimer's disease: clinical assessment and metabolic correlates. Am J Geriatr Psychiatry 25(7):730–740. https://doi.org/10.1016/j.jagp.2017.02.002

32. John GH, Kohavi R, Pfleger K (1994) Irrelevant features and the subset selection problem. In: Cohen WW, Hirsh H (eds) Machine Learning Proceedings 1994. Morgan Kaufmann, pp 121–129

33. Koller D, Friedman N Probabilistic graphical models. Principles and Techniques. The MIT Press. https://books.google.co.in/books?id=7dzpHCHzNQ4C. Accessed 23 Oct 2021

34. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI (2015) Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 13:8–17

35. Luce L (2019) Democratization and impacts of ai. Apress, Berkeley, pp 185–195. https://doi.org/10.1007/978-1-4842-3931-5_12

36. Luo G (2016) A review of automatic selection methods for machine learning algorithms and hyper-parameter values. Network Modeling Analysis in Health Informatics and Bioinformatics, 5. https://doi.org/10.1007/s13721-016-0125-6

37. Matías-Guiu J A, N. C-MM, Valles-Salgado M, Rognoni T, Galán L, Moreno-Ramos T, Carreras JL, Matías-Guiu J (2019) Inhibition impairment in frontotemporal dementia, amyotrophic lateral sclerosis, and Alzheimer's disease: clinical assessment and metabolic correlates. Brain Imaging and Behavior 13 (3):651659. https://doi.org/10.1007/s11682-018-9891-3

38. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, et al. (2011) The diagnosis of dementia due to Alzheimer's disease: recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & Dementia 7:263–269

39. Nordberg A, Rinne JO, Kadir A, Långström B (2010) The use of pet in Alzheimer disease. Nat Rev Neurol 6:78–87

40. Olazarán J, Mouronte P, Bermejo F (2005) [Clinical validity of two scales of instrumental activities in Alzheimer's disease]. Neurologia 20:395–401

41. Ossenkoppele R, Singleton EH, Groot C, Dijkstra AA, Eikelboom WS, Seeley WW, Miller B, Laforce RJ, Scheltens P, Papma JM, Rabinovici GD, Pijnenburg YAL (2022) Research criteria for the behavioral variant of Alzheimer disease: a systematic review and meta-analysis. JAMA Neurol 79:48–60. https://doi.org/10.1001/jamaneurol.2021.4417

42. Pagán J, Risco-Martín JL, Moya JM, Ayala JL (2016) Grammatical evolutionary techniques for prompt migraine prediction. In: Proceedings of the Genetic and Evolutionary Computation Conference 2016, Association for Computing Machinery, New York, NY, USA. p. 973980. https://doi.org/10.1145/2908812.2908897

43. Pena D, Barman A, Suescun J, Jiang X, Schiess MC, Giancardo L, The Alzheimer's Disease Neuroimaging Initiative (2019) Quantifying neurodegenerative progression with deepsymnet, an end-to-end data-driven approach. Front Neurosci 13:1053. https://doi.org/10.3389/fnins.2019.01053

44. Puente-Castro A, Fernandez-Blanco E, Pazos A, Munteanu CR (2020) Automatic assessment of Alzheimer's disease diagnosis based on deep learning techniques. Comput Biol Med 120:103764. https://doi.org/10.1016/j.compbiomed.2020.103764

45. Rajkomar A, Dean J, Kohane I (2019) Machine learning in medicine. N Engl J Med 380 (14):1347–1358

46. Rascovsky K, Hodges JR, Knopman D, Mendez MF, Kramer JH, Neuhaus J, van Swieten JC, Seelaar H, Dopper EG, Onyike CU, Hillis AE, Josephs KA, Boeve BF, Kertesz A, Seeley WW, Rankin KP, Johnson JK, Gorno-Tempini ML, Rosen H, Prioleau-Latham CE, Lee A, Kipps CM, Lillo P, Piguet O, Rohrer JD, Rossor MN, Warren JD, Fox NC, Galasko D, Salmon DP, Black SE, Mesulam M, Weintraub S, Dickerson BC, Diehl-Schmid J, Pasquier F, Deramecourt V, Lebert F, Pijnenburg Y, Chow TW, Manes F, Grafman J, Cappa SF, Freedman M, Grossman M, Miller BL (2011) Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain 134:2456–2477

47. Ryan C, O'Neill M (1998) Grammatical evolution: a steady state approach. In: Koza JR (ed) Late Breaking Papers at the Genetic Programming 1998 Conference, Stanford University Bookstore, University of Wisconsin, Madison, Wisconsin, USA. pp. 180-185.

48. Saeys Y, Inza I, Larrañaga P (2007) A review of feature selection techniques in bioinformatics. Bioinformatics 23:2507–2517. https://doi.org/10.1093/bioinformatics/btm344

49. Schreiber J (2017) Pomegranate: fast and flexible probabilistic modeling in Python. J Mach Learn Res 18:5992–5997

50. Stekhoven DJ, Bühlmann P (2012) Missforestnon-parametric missing value imputation for mixed-type data. Bioinformatics 28:112–118

51. Stern Y (2021) How can cognitive reserve promote cognitive and neurobehavioral health?. Arch Clin Neuropsychol 36:1291–1295. https://doi.org/10.1093/arclin/acab049

52. Tăuţan A-M, Ionescu B, Santarnecchi E (2021) Artificial intelligence in neurodegenerative diseases: a review of available tools with a focus on machine learning techniques. Artif Intell Med 117:102081

53. Tracy JM, Özkanca Y, Atkins DC, Hosseini Ghomi R (2020) Investigating voice as a biomarker: Deep phenotyping methods for early detection of parkinson's disease. J Biomed Inform 104:103362. https://doi.org/10.1016/j.jbi.2019.103362

54. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. NeuroImage 15:273–289. https://doi.org/10.1006/nimg.2001.0978

55. Vehí J, Contreras I, Oviedo S, Biagi L, Bertachi A (2020) Prediction and prevention of hypoglycaemic events in type-1 diabetic patients using machine learning. Health Inf J 26:703–718. https://doi.org/10.1177/1460458219850682 PMID: 31195880

56. Wolpert DH (1992) Stacked generalization. Neural Netw 5:241–259. https://doi.org/10.1016/S0893-6080(05)80023-1

57. Wu J, Azarm S (2001) Metrics for quality assessment of a multiobjective design optimization solution set. J Mech Des 123:18–25

58. Xue B, Zhang M, Browne WN, Yao X (2016) A survey on evolutionary computation approaches to feature selection. IEEE Trans Evol Comput 20(4):606–626

59. Zhang H, Sun G (2002) Feature selection using tabu search method. Pattern Recogn 35:701–711. https://doi.org/10.1016/S0031-3203(01)00046-2

Fernando García-Gutierrez is a MsC in Bioinformatics pursuing his PhD in Computer Science. His research interests are in the area of machine learning application to the diagnosis and prognosis of neurodegenerative diseases.

Josefa Díaz-Álvarez is Assistant Professor in Computer Architecture at Extremadura University. PhD in Computer Engineering in Complutense University of Madrid. She is interested in bioinspired algorithms, and computing applied to bioengineering.

Jordi A. Matias-Guiu is a neurologist in the Neurology Service of the Hospital Clínico San Carlos, Madrid. PhD in Medicine, Complutense University of Madrid. His research activity is focused on neuropsychological assessments, neuroimaging, and non-invasive neuromodulation.

Vanesa Pytel is a neurologist in ACE Alzheimer Center, Barcelona. PhD in Medicine, Complutense University of Madrid. Her research interests are in the field of transcranial stimulation applied to neurodegenerative diseases and genetic profiling.

Jorge Matías-Guiu is a Professor of Neurology, Universidad Complutense de Madrid. Director of the Neuroscience Institute of Hospital Clinico San Carlos, and Head of Neurology Service.

María Nieves Cabrera-Martín is a specialist in Nuclear Medicine in the Hospital Clínico San Carlos, Madrid. PhD in Medicine, Complutense University of Madrid. Research interests on the validation of 18-FDG PET for the diagnosis of neurodegenerative diseases.

José L. Ayala is a Professor in Computer Architecture and Automation in Complutense University of Madrid. PhD in Computer and Electrical Engineering, Technical University of Madrid. Research interests in computing applied to bioengineering.