



Medical Informatics Platform (MIP): A Pilot Study Across Clinical Italian Cohorts

Alberto Redolfi^{1*}, Silvia De Francesco^{1,2}, Fulvia Palesi^{3,4}, Samantha Galluzzi², Cristina Muscio⁵, Gloria Castellazzi^{4,6,7}, Pietro Tiraboschi⁵, Giovanni Savini⁴, Anna Nigri⁸, Gabriella Bottini⁹, Maria Grazia Bruzzone⁸, Matteo Cotta Ramusino^{4,10}, Stefania Ferraro⁸, Claudia A. M. Gandini Wheeler-Kingshott^{3,4,6}, Fabrizio Tagliavini⁵, Giovanni B. Frisoni^{2,10}, Philippe Ryvlin¹¹, Jean-François Demonet¹¹, Ferath Kherif¹¹, Stefano F. Cappa^{4,12} and Egidio D'Angelo^{3,4} for the ADNI – HBP – Italian Network of Neuroscience Neurorehabilitation (RIN) Initiatives

OPEN ACCESS

Edited by:

Roberto Monastero,
University of Palermo, Italy

Reviewed by:

Pieter Visser,
Maastricht University, Netherlands
Antonio Cerasa,
Institute for Biomedical Research and
Innovation (CNR), Italy

*Correspondence:

Alberto Redolfi
aredolfi@fatebenefratelli.eu

Specialty section:

This article was submitted to
Dementia and Neurodegenerative
Diseases,
a section of the journal
Frontiers in Neurology

Received: 28 February 2020

Accepted: 04 August 2020

Published: 23 September 2020

Citation:

Redolfi A, De Francesco S, Palesi F, Galluzzi S, Muscio C, Castellazzi G, Tiraboschi P, Savini G, Nigri A, Bottini G, Bruzzone MG, Ramusino MC, Ferraro S, Gandini Wheeler-Kingshott CAM, Tagliavini F, Frisoni GB, Ryvlin P, Demonet J-F, Kherif F, Cappa SF and D'Angelo E (2020) Medical Informatics Platform (MIP): A Pilot Study Across Clinical Italian Cohorts. *Front. Neurol.* 11:1021. doi: 10.3389/fneur.2020.01021

¹ Laboratory of Neuroinformatics, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy, ² Laboratory of Alzheimer's Neuroimaging and Epidemiology - LANE, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy, ³ Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy, ⁴ IRCCS Mondino Foundation, Pavia, Italy, ⁵ Division of Neurology V/Neuropathology, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy, ⁶ NMR Research Unit, Queen Square MS Center, Department of Neuroinflammation, UCL Institute of Neurology, London, United Kingdom, ⁷ Department of Computer, Electrical and Biomedical Engineering, University of Pavia, Pavia, Italy, ⁸ Department of Neuroradiology, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy, ⁹ Neuropsychology Center, ASST Grande Ospedale Metropolitano Niguarda, Milan, Italy, ¹⁰ Memory Clinic and LANVIE - Laboratory of Neuroimaging of Aging, University Hospitals and University of Geneva, Geneva, Switzerland, ¹¹ Department of Clinical Neurosciences, Leenaards Memory Center, Center Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland, ¹² University School of Advanced Studies, Pavia, Italy

Introduction: With the shift of research focus to personalized medicine in Alzheimer's Dementia (AD), there is an urgent need for tools that are capable of quantifying a patient's risk using diagnostic biomarkers. The Medical Informatics Platform (MIP) is a distributed e-infrastructure federating large amounts of data coupled with machine-learning (ML) algorithms and statistical models to define the biological signature of the disease. The present study assessed (i) the accuracy of two ML algorithms, i.e., supervised Gradient Boosting (GB) and semi-supervised 3C strategy (Categorize, Cluster, Classify—CCC) implemented in the MIP and (ii) their contribution over the standard diagnostic workup.

Methods: We examined individuals coming from the MIP installed across 3 Italian memory clinics, including subjects with Normal Cognition (CN, $n = 432$), Mild Cognitive Impairment (MCI, $n = 456$), and AD ($n = 451$). The GB classifier was applied to best discriminate the three diagnostic classes in 1,339 subjects, and the CCC strategy was used to refine the classical disease categories. Four dementia experts provided their diagnostic confidence (DC) of MCI conversion on an independent cohort of 38 patients. DC was based on clinical, neuropsychological, CSF, and structural MRI information and again with addition of the outcome from the MIP tools.

Results: The GB algorithm provided a classification accuracy of 85% in a nested 10-fold cross-validation for CN vs. MCI vs. AD discrimination. Accuracy increased to 95% in the holdout validation, with the omission of each Italian clinical cohort out in turn. CCC

identified five homogeneous clusters of subjects and 36 biomarkers that represented the disease fingerprint. In the DC assessment, CCC defined six clusters in the MCI population used to train the algorithm and 29 biomarkers to improve patients staging. GB and CCC showed a significant impact, evaluated as +5.99% of increment on physicians' DC. The influence of MIP on DC was rated from "slight" to "significant" in 80% of the cases.

Discussion: GB provided fair results in classification of CN, MCI, and AD. CCC identified homogeneous and promising classes of subjects via its semi-supervised approach. We measured the effect of the MIP on the physician's DC. Our results pave the way for the establishment of a new paradigm for ML discrimination of patients who will or will not convert to AD, a clinical priority for neurology.

Keywords: Alzheimer's Dementia (AD), biomarkers, diagnostic confidence, Medical Informatics Platform (MIP), disease signature

INTRODUCTION

The International Working Group (IWG) criteria (1, 2) define Alzheimer's Dementia (AD) as a construct based on the combination of brain amyloidosis and tauopathy (3–6). However, physicians in real-world practice primarily rely on the patient's clinical picture. Neurological diseases are classified into internationally recognized catalogs (ICDM, DSM-V) that are essentially based on phenotyping backed up by ancillary investigations.

Neuroscience seems locked in a symptom-based diagnostic paradigm moving forward without a wide-ranging biological classification scheme for brain diseases (7). To reach a complete understanding of the brain, neuroscientists cannot just focus on one level of brain organization (8), regardless of how well it may be determined at that specific level (9). An integrated, holistic, multimodal, and multilevel analysis is required. This approach could optimize the effectiveness of disease prevention and treatment and minimize side effects by considering the specific makeup of genetic information, *in vivo* biomarkers (e.g., biological or imaging-derived), and phenotypic characterization to move toward the so-called personalized medicine.

However, if neuroscience desires to understand the brain across all scales and levels of organization, it needs robust and efficient computational e-infrastructures to work together (10, 11). Computational power is mandatory to solve complex brain simulations and forecasts as quickly as possible. To cope with this need, many platforms were developed in Europe and North America in the last decade to fill the gap between data collection and information extraction (12). Neuroscience e-infrastructures, such as Laboratory Of Neuro Imaging (LONI) (13), neuGRID (14–16), C-Brain (17), and EMIF (18, 19), offer access to large databases, neuroimaging algorithms, extended computational resources, and statistical tools. More recently, the Human Brain Project (HBP), funded in 2012 by the European Commission to build an open e-infrastructure, started to provide supervised/unsupervised machine learning (ML) tools for neuroinformatics, brain simulation, neuromorphic computing, and medical informatics. Within this overarching project, the Medical Informatic Platform (MIP) was developed

to allow hospitals and research centers to share medical data (20). Notably, patient records in European hospitals represent an enormous source of data (21) that are waiting to be processed using data mining algorithms and mathematical models to extract meaningful and potentially hidden information. Therefore, this vast amount of available data and emerging powerful algorithms sets the stage for a paradigm shift from a pure-hypothesis-driven medicine, as adopted currently, to a data-led objective classification strategy. This shift is the overarching aim of the MIP to find a way to combine ML tools with big data repositories already collected in hospitals to search for disease signatures (22).

The Italian Network of Neuroscience and Neurorehabilitation (RIN) (23) was founded in 2017 by the Italian Ministry of Health, and it represents the ideal starting point to test the MIP platform and validate the "disease signature" concept against clinical practice. The RIN initiative (i) encourages collaboration among the Italian Research Hospitals (IRCCS), (ii) facilitates the spread of information on clinical/scientific activities, and (iii) promotes collaborative actions at international level. Clinical, instrumental, and molecular characterization of patients represent the basis of RIN studies to identify advanced early diagnostic biomarkers, therapeutic targets, and innovative intervention strategies. Two IRCCS (IRCCS Fatebenefratelli and IRCCS Besta) of the RIN network and another external Italian hospital (Niguarda Ca' Granda) in association with the NeuroImaging lab of the Center for Health Technologies (CHT—University of Pavia) configured MIP platforms locally and participated in this study.

Data availability is as important as data comparability. Recent advances in computer science and the widespread application of big data mining have demonstrated that meaningful insights may be obtained from heterogeneous, noisy, non-standardized data (24). The variety of normal ranges is generally handled with normalization and z-scoring (25). Dichotomization is used to define abnormality when validated cut-points for abnormal scores were developed. "Messy" data may be cleaned up via smoothing, which is a technique long used in brain imaging with excellent effect (26). Missing data may be imputed using specialized interpolations, also used effectively in the averaging of brain images (27). Scaling approaches are efficiently used in

case of systematic differences between cohorts for scores on the same variable (28). Strategies for coping with non-standardized data are already in place, and MIP takes advantage of all of these strategies to gather large numbers.

This study measured the performance of a supervised MIP Gradient Boosting (GB) classifier in discriminating subjects in one of the three broad categories originally attributed by clinical experts: Cognitive Normal (CN), Mild Cognitive Impairment (MCI), and Alzheimer's Dementia (AD). The initial diagnosis of the subjects was based on clinical and neuropsychological testing (6, 29, 30), because the clinical judgment is still the gold standard for the syndromic diagnosis (31). Then, we introduced the idea of classifying subjects into more than three diagnostic classes using a semi-supervised approach that exploited the so-called 3C strategy (Categorization, Clustering, Classification—CCC) for refinement of the common disease diagnoses, beyond what the eye can see (32). Both tools were chosen for their high degree of maturity in the MIP development factory.

Further, we investigated the perception of the clinical utility of the MIP by interviewing four clinical experts at the three participating leading Italian dementia centers to provide an assessment of diagnostic confidence (DC). We then compared the clinicians' DC values based on the MIP information with those of traditional workup information in patients with MCI.

Clinical DC truly determines patient management, and a quantifiable measure of how the ML tools of the MIP were perceived by the clinicians and how they could potentially be incorporated in a clinical context would be highly beneficial.

E-infrastructures, which may undergo substantial reshaping in the near future, involving ML algorithms and analyses of big data, may offer a solution to the aforementioned disconnection between the biological and clinical levels of disease description in the long run. If the MIP paradigm is successful, it will provide opportunities to design revolutionary *in silico* experiments to examine and elucidate the mechanisms of brain diseases in ways that were impossible until a few years ago.

MATERIALS AND METHODS

Study Design

Our study was structured in two main parts: (1) a “group analysis” to assess the performances of two MIP ML algorithms to classify subjects in the dementia spectrum and to identify informative disease signatures and (2) a “single-case analysis” to measure the clinicians' perceived impact of the MIP tools and test the change in their DC to identify MCI that could convert to AD within 2 years. The same algorithms were used in both phases, i.e., GB (33) and CCC (34). **Supplementary Figure 1** shows the workflows of the study.

Medical Informatics Platforms

We installed and configured MIPs in three leading Italian dementia centers. The MIP platform is an e-infrastructure for data federation and big-data analysis. It was primarily developed by Center Hospitalier Universitaire Vaudois (CHUV) in the context of the HBP (<https://www.humanbrainproject.eu/>).

MIP enables access and analyses of anonymized medical data that are currently locked in hospitals without moving the raw data from the servers where they reside or infringe on patient privacy. End-users cannot explore the local database of each hospital. Only aggregated results or features are shared outside of the hospital.

The MIP is organized into two main parts (see **Supplementary Figure 2**). The MIP-Local, where only the data coordinator and associated staff can access the pseudonymized data set, and the MIP Federated, where different hospitals are connected and the end user can query and run analyses on the federated fully anonymized data sets (35). The platform provides algorithms for advanced statistical analyses, feature extraction, and predictive models via data mining and ML tools. The MIP infrastructure overall is instrumental to the identification of the biological changes associated with AD and opens new possibilities for early diagnosis by discovering otherwise unseen disease signatures.

Data

For the “group analysis,” we used data from five different cohorts. Two of the cohorts were international research data sets, such as: The European DTI Study on Dementia (EDSD) (36) and the Alzheimer's Disease Neuroimaging Initiative (ADNI) (37) that are constituent parts of each MIP platform. The other three cohorts were Italian clinical cohorts collected from IRCCS Fatebenefratelli (FBB), IRCCS Carlo Besta (Besta), and Niguarda hospital (CHT-Niguarda). The local medical ethics committee approved each study. Participants provided written informed consent at the time of inclusion in the study for the sharing of cognitive data, fluid samples, and MRI scans. The total number of subjects originally collected was 2105 (559 AD, 847 MCI, and 699 CN). Inclusion and exclusion criteria are reported in **Supplementary Table 1**. The sample size was reduced to 1339 (451 AD, 456 MCI, 432 CN) to obtain data-set matching and prevent class imbalance and after visual quality control of the available T1D scans and the relative imaging biomarker segmentations of experienced researchers (AR, SD).

For the “single-case analysis,” we considered 198 MCI subjects from ADNI (118 MCI stable and 80 MCI converters to AD) with a 24-month clinical follow up as a training set. The independent subjects' data for the clinical testing came from the FBB-HUG-2014 Piramal study (38) (alias, testing set) led by the University of Geneva, with 38 MCI possibly due to AD subjects (6).

The characteristics of all the data cohorts used in the present study are reported in **Table 1**. Any variable with more than 40% of missing data was discarded. All the cohorts considered in the present study exposed 144 variables and included metrics in the following categories: socio-demographical, neuropsychological (including tests for verbal memory, attention, and language), volumetric (from MRI), genetic, and cerebrospinal fluid (CSF) information. In particular, Mini-Mental State Examination (MMSE) scores were available for all of the subjects considered. Surrogate imaging biomarkers were extracted from 3D T1-weighted (3DT1) MRIs. Imaging was acquired according to local Magnetization Prepared Rapid Acquisition Gradient Echo (MPRAGE) or Inversion

TABLE 1 | Characteristics of the data sets.

Data Set	Types and availability	Description	Categories
ADNI—Alzheimer's Disease Neuroimaging Initiative	Research data set available on every MIP web-server interface (https://services.humanbrainproject.eu/oicd/login)	A longitudinal multicenter study designed to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD. ADNI was originally launched in 2003 as a public-private partnership; its primary goal has been to test whether MRI, PET, biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of MCI and Alzheimer's disease. It has targeted participants with AD ($n = 222$), MCI ($n = 576$), and CN ($n = 268$).	CN MCI AD
EDSD—European DTI Study on Dementia	Research data set available on every MIP web-server interface (https://services.humanbrainproject.eu/oicd/login)	A framework of nine European centers: Amsterdam (Netherlands), Brescia (Italy), Dublin (Ireland), Frankfurt (Germany), Freiburg (Germany), Milano (Italy), Mainz (Germany), Munich (Germany), and Rostock (Germany). It is a cross-sectional multicenter study characterized by 474 volumetric MRI T1-weighted scans with socio-demographic, clinical, genetic, and biological variables.	CN MCI AD
FBF	Clinical data set collected by the IRCCS Fatebenefratelli and available via MIP-Local and MIP-Federated	A cross-sectional collection including data from 1,784 patients enrolled in Brescia (Italy) and nearby areas. The data set contains socio-demographic, clinical, genotype, bio-specimen information, and MRI T1-weighted images	CN MCI AD Other
Besta	Clinical data set collected by the IRCCS Carlo Besta hospital and available via MIP-Local	A cross-sectional collection of 105 patients clinically characterized using standard neuropsychological evaluation. Volumetric MRI scans at 3T are available for all patients. CSF and DNA were collected in most cases to investigate biomarkers (e.g., T-TAU, P-TAU, AB1-42) and genotyping.	CN MCI AD Other
CHT Niguarda	Clinical data set collected by the Niguarda hospital and available via MIP-Local	A cross-sectional collection of 171 subjects. Most of subjects underwent clinical/neuropsychological assessments and volumetric MRI acquisitions.	CN MCI AD Other
Piramal	Independent external research data set not federated with the MIP infrastructure	An international, multicenter, phase 4 study where a fraction of the total patients enrolled ($n = 71$) were characterized using clinical, neuropsychological assessments, MRI (consistent with ADNI acquisition protocol), amyloid-PET, and CSF biomarkers.	MCI

CSF, Cerebrospinal Fluid; AB1-42, amyloid beta; T-TAU, total Tau; P-TAU, phosphorylated Tau; MRI, Magnetic Resonance Imaging; CN, Cognitive Normal; MCI, Mild Cognitive Impairment; AD, Alzheimer's Disease; Other, atypical form of dementia in the clinical data sets of each hospital - these subjects were excluded from our study.

Recovery SPOilt Gradient echo (IRSPGR) acquisition protocols. Volumetric scans were processed using the Neuromorphometrics pipeline (Neuromorphometrics, Inc., Somerville, MA) (39, 40) integrated in the MIP environment and the volumes normalized to the total intracranial volume (TIV) computed with Statistical Parametric Mapping (SPM12) considering a reference intracranial volume of 1,409 ml (41). APOE genotyping data from local genetic analyses were available for 65% of the selected individuals. CSF biomarkers were obtained using different assays across different cohorts, i.e., the Multiplex xMAP Luminex platform with Innogenetics immunoassay kit-based reagents (42) for ADNI subjects and Enzyme-Linked Immunosorbent Assay (ELISA) (43, 44) for subjects from all other cohorts, which led to different CSF biomarker distributions. To tackle this issue, biomarkers from clinical cohorts, ADNI, and EDSD were Z-scored based on the normative data specific for each cohort. All of the variables considered in the MIP were adjusted by age. A linear detrending algorithm based on age-related changes in the CN group was adopted (45) as age correction method. The adopted approach fitted a generalized linear model (GLM) for each variable and age, and the age-related changes were modeled as linear drift in the CN group only. The regression

coefficient of the resulting GLM model was used to remove the age-related changes from all individuals to obtain corrected values. The assumption for the age correction method is that the age-related changes in the CN group are due to aging while the age-related changes in the MCI and AD groups includes disease-related changes.

Categorization of the MIP Variables

Categorization in three main sets was performed to give a logical organization of the 144 variables specified in the input to the MIP supervised and semi-supervised algorithms.

The first category included disease diagnosis (as assigned in the original cohort) and was used in the evaluation of the supervised GB classification performances. The diagnosis feature had three levels: CN, MCI, and AD.

The second category was represented by 12 validated MRI biomarkers that very well-described the brain neurodegeneration (6). These markers were the volumes of the following structures: left/right hippocampus, left/right amygdalae, left/right anterior cingulate gyrus, left/right middle cingulate gyrus, left/right posterior cingulate gyrus, and 3rd and 4th ventricles. Both GB and CCC used this set. Furthermore, the second category was

used in the CCC algorithm to define homogeneous clusters of subjects.

The third category included all the other measurements among those of the MIP we used (for a complete list of the categorization refers to **Supplementary Table 2**) for which we knew had, or potentially could have, a predictive value for disease risks or severity. A total of 131 markers were in this category, such as sociodemographic variables, MMSE, CSF proteins, APOE $\epsilon 4$ genotyping, and all other less common Neuromorphometric derived imaging variables. GB and CCC used this category. The most informative features in the third category, in addition to those in the second category, were explicitly used to define the disease fingerprint of the CCC algorithm (see **Supplementary Table 3**).

Gradient Boosting

GB is a popular ML technique to solve classification problems. GB produces a prediction that exploits an ensemble of weak estimators (i.e., decision trees). GBs are used extensively in neuroimaging as because they predict outcomes with high accuracy and possess the ability to model diverse and high-dimensional data (46, 47). The main advantages of using GB are its capacity to handle variables of mixed types and its inner robustness to outliers.

In the “group analysis,” we constructed a classifier to separate patients into the following three groups: CN, MCI, and AD, i.e., the values of the first category. We performed feature relevance evaluation using a tree-based approach with a nested fold cross-validation design. The nested cross-validation consists of an inner loop for model building and parameter estimation and an outer loop for model testing. Consequently, the data set was divided into two parts: a training plus validation subset and a test subset. In the inner loop, GB models were trained with varying GB hyper-parameters (e.g., learning rate, number of estimators, maximum number of features, minimum samples split, maximum depth) based on a grid-search strategy. The validation set was used to determine the GB hyper-parameters over the grid of possible values. The performance of the resulting model, with optimized GB hyper-parameters and features, was subsequently evaluated on the test set in the outer loop. For this outer loop, we used a stratified 10-fold cross-validation scheme to divide the data into 10 equally sized parts. Nine of these parts were dedicated to the training/validation set, and one part was the test set. The 10 parts were permuted in each iteration of the outer loop so that each one was used for testing once. Finally, the GB results were averaged over the 10-folds to estimate the predictive power of the proposed model on the entire data set. To further test the GB generalizability and flexibility performances, we performed many holdout validation assessments between research and clinical data sets as well as on each clinical data set.

Before the “single-case analysis” was performed on PIRAMAL subjects, a nested 5-fold cross-validation was used to train and test the GB on the subgroup of ADNI MCI patients. Accuracy, precision, and recall metrics were computed to assess performance.

The GB algorithm, as implemented in the open-source python Scikit-learn library (version 0.22.1), was used to perform classification (48).

CCC Algorithm

The CCC algorithm is a semi-unsupervised ML tool developed by Tel Aviv University (32, 34). CCC was used to obtain a homogeneous subjects’ clustering and to identify potential combinations of biomarkers for a deep characterization of the disease.

In the first step, the algorithm considered the so-called second category variables that we defined earlier, which was composed of 12 imaging features, plus the clinical diagnosis to identify the optimal number of clusters present in the patient’s data set. A Random Forest (RF) selection method ranked the weight of each feature, and the number of clusters was derived using Gap statistic (49). Then, the clustering PAM (Partitioning Around Medoids, also known as K-medoids) algorithm (50) was applied to label each subject into one cluster. To discuss the meaning of the created classes, we cross-classified the clusters generated with the original diagnosis and demographic and clinical variables. Finally, the still unused variables of the third category were exploited as potential features to define and expand the disease fingerprints of the created clusters. In particular, the CCC algorithm selected a subset among the most promising of the 131 third category variables. To do this, a feature selection process was performed via the RF mechanism. RF assigned a weight of importance to each of the third category feature (51), and the most informative were selected. In this way, CCC identified markers that were useful to define the final disease fingerprint.

Further, the CCC algorithm, prying on the selected informative third category variables and hierarchical decision trees with out-of-bag validation (52), generated a matching matrix that represented the ability to correctly classify subjects according to the K-medoids clusters previously identified.

Parallel coordinate plots were used to allow comparison of the derived fingerprints, including confidence intervals. Data were normalized by subtracting the minimum and dividing by the maximum of all observations. This method allowed comparisons of variables of different scales and keep preserved the shape of the different distributions.

The CCC is an MIP semi-unsupervised strategy and is based on the “randomForest,” “cluster,” “rpart,” “psych,” and “ggparci” R packages.

Clinical Impact

The “single-case analysis” we performed tested and measured how physicians are influenced in their DC by the MIP information. The clinical question was to verify the perceived usefulness of the MIP to classify whether an MCI subject would convert to AD or would remain stable in the next 2 years. Physicians were also asked to express the etiological causes of the disease. The 4 physicians (SG, GB, PT, SC) had a long experience with the diagnostic use of AD biomarkers and were aware of the most recent research diagnostic criteria (1, 2, 5) for AD. To provide useful information to the clinicians, the MIP algorithms were trained on 118 MCI stable and 80 MCI

converters to AD from ADNI. Thirty-eight MCI subjects from Piramal were analyzed as multiple single cases. Piramal 3DT1 scans were used to extract features using the Neuromorphometric pipeline, and all the data were post-processed with both GB and CCC, while the clinical assessments were collected using an *ad hoc* web-based questionnaire.

The physicians interviewed did not have specific constraints about the diagnostic etiologies. All etiological diagnoses were grouped ex-post into:

- (i) Pathophysiologically AD-related: MCI due to AD
- (ii) Not pathophysiologically AD-related: MCI not due to AD, MCI due to frontotemporal dementia (FTD); MCI due to Vascular Disease (VD); MCI associated with psychiatric disorder; Suspected Non-Amyloid Pathology (SNAP); and Normal Aging (**Supplementary Table 4**).

The four physicians provided their initial estimate of DC. The following available information for each Piramal subject, describing the initial physicians' evaluations, was used: (1) age, sex, education, and other socio-demographic information; (2) neurological examination; (3) MMSE score; (4) neuropsychological assessment of long-term memory, executive functions, language domains expressed as both raw and equivalent scores (53); (5) MRI visual assessment performed by an expert neuro-radiologist; and (6) amyloid beta (AB1-42), total Tau (T-TAU), and phosphorylated Tau (P-TAU) CSF levels.

Then, the four physicians were asked to reassess the same subjects by taking into consideration the report from the MIP, which included the following factors: (i) the density function distributions built from the Neuromorphometrics analysis of established imaging biomarkers (i.e., hippocampus, cingulate, entorhinal, parahippocampal, and superior temporal gyrus) showing three standard deviations (**Supplementary Figure 3**); (ii) the CCC K-medoids clustering (**Figure 2B**); and (iii) GB classification results, which ranged from a stable stage to a very probable conversion stage (**Supplementary Figure 4**), depicted via a partial dependence plot (PDP). The four physicians provided their final estimate of DC on a structured scale (ranging from 0 to 100%). Clinicians finally reported whether the ML information of the MIP affected their DC assessment for each subject using a four-level Likert scale (possible answers: YES significantly; YES somewhat; YES slightly; NO not at all).

Statistical Analyses

As far as “group analysis” is concerned, MANOVA was initially performed in R (version 3.5.1) to verify the comparability of the five data sets in the three diagnostic classes (CN, MCI, and AD).

Kruskal–Wallis and Fisher's exact *post hoc* tests ($\alpha = 0.05$) were, respectively, used to test continuous or binary markers differences between the three diagnostic groups.

With regard to “single-case analysis”, Fleiss' Kappa inter-physicians reproducibility on etiological diagnosis, with and without MIP information available, was performed in R (v3.5.1) using the “dplyr” and “irr” packages. We used the python Scikit-learn library (version 0.22.1) for the “single-case analysis” to generate the PDP of GB showing each Piramal subject plotted in

a 2D probability space. PDP depicted the probability of being an MCI converter using a canonical 2D space.

RESULTS

Group Definition

To test the comparability, we performed a MANOVA test on the three diagnostic classes for the five data sets (see **Supplementary Table 5**). All data set comparisons performed were statistically not dissimilar for the same diagnostic class ($p > 0.05$). Our group included 1,339 subjects, who were stratified as follows: 432 CN (195 FBF, 28 Besta, 3 CHT-Niguarda, 100 ADNI, 106 EDS); 456 MCI (103 FBF, 16 Besta, 18 CHT-Niguarda, 269 ADNI, 50 EDS); 451 AD (200 FBF, 15 Besta, 135 ADNI, 101 EDS).

Demographic and clinical information are presented in **Table 2**. The AD subjects were older, were less educated, and had lower MMSE compared to the other groups. We observed a female gender preponderance in the CN and AD groups. AD subjects were more often APOE $\epsilon 4$ carriers than MCI and CN subjects. Within the CN group, when quantified, 9% of subjects had abnormal AB1-42 CSF marker, and 82 and 97% of subjects had this abnormal marker in the MCI group and AD group, respectively.

For the Piramal data set, 53% of the 38 MCI subjects were male, the mean age was 71.03 ± 7.72 years, and the mean raw MMSE score was 26.3 ± 2.14 . The overall frequencies of the considered clinical measures and biomarkers assessed are reported in **Table 3**.

Gradient Boosting Results

In the “group analysis,” the most informative features selected by the GB classifier in order of importance were AB1-42, T-TAU, P-TAU, and MMSE. Other important, albeit less informative features, selected across the groups were left hippocampus, left amygdala, and 4th ventricle. The complete rank of features is reported in **Supplementary Table 6**.

In the nested 10-fold cross-validation that combined all of the informative features, GB resulted in an accuracy of $85 \pm 4\%$ in classifying CN, MCI, and AD subjects. Generally, exclusion of the CSF variables led to a deterioration of the GB performance. The results of GB in holdout validations, which was used to test independently each clinical data set, resulted in overall better performances, with an accuracy that ranged from 91 to 95%. **Table 4** shows the Accuracy, Precision and Recall metrics. GB revealed high precision and high recall in the “group analysis,” which related to a low false-positive and low false-negative rates, respectively.

In preparation of the “single-case analysis,” we assessed the GB performances for identifying MCI converters to AD using a nested 5 cross-fold validation strategy performed on 198 MCI ADNI patients. These patients were followed longitudinally for 24 months and expert ADNI neurologists performed their clinical assessments. The final assessment of ADNI physicians represented our ground truth and yielded an accuracy of $62 \pm 1\%$. This base allowed us to fine-tune the GB hyperparameters in

TABLE 2 | Group characteristics.

Characteristics	Cognitively normal - CN; n = 432	Mild cognitive impairment - MCI; n = 456	Alzheimer disease - AD; n = 451	P
Age—years	65.47 (9.31)	71.18 (8.08)	73.70 (7.97)	<0.001 $\text{c}, \text{¥}, \text{z}$
Gender—M/F	172/260	228/228	168/283	<0.001 c, z
Education—year	12.60 (4.36)	13.53 (4.40)	10.64 (4.75)	<0.001 $\text{c}, \text{¥}, \text{z}$
MMSE—raw score	27.91 (1.27)	26.33 (2.01)	20.54 (4.25)	<0.001 $\text{c}, \text{¥}, \text{z}$
% APOE ϵ 4 carriers				
—/—	80%	65%	25%	
—/+	18%	30%	64%	
+ /+	2%	5%	11%	
% AB1-42 abnormal	9% [23%]	82% [68%]	97% [38%]	<0.001 $\text{c}, \text{¥}, \text{z}$
% T-TAU abnormal	4% [23%]	17% [69%]	90% [38%]	<0.001 $\text{c}, \text{¥}, \text{z}$
% P-TAU abnormal	4% [23%]	18% [69%]	62% [37%]	<0.001 $\text{c}, \text{¥}, \text{z}$

Data are expressed as the mean values. Standard deviation is reported in round brackets. The overall frequencies of the CSF biomarkers in each group are reported in square brackets. AB1-42, amyloid beta; T-TAU, total Tau; P-TAU, phosphorylated Tau; n, numerosity; P, significance on Fisher's exact post hoc test or Kruskal–Wallis multiple comparison test with the Benjamini–Hochberg method (c , Significant difference between CTR and MCI; ¥ , Significant difference between CTR and AD; z , Significant difference between MCI and AD).

TABLE 3 | Descriptive statistics of the clinical variables and biomarkers of the Piramal patients.

Characteristics	Piramal mild cognitive impairment group	
	Biomarker frequency	Mean (SD)
Age—years	100%	71.03 (7.72)
Gender—M/F		20/18
Education—year	100%	11.29 (4.63)
MMSE—raw score	100%	26.3 (2.14)
Logical Memory—EQS	100%	0.89 (1.11)
AVLT Immediate—EQS	100%	2.24 (1.42)
AVLT Delayed—EQS	100%	1.53 (1.39)
Digit Span Forward—EQS	100%	3.08 (1.12)
Digit Span Backward—EQS	53%	2.8 (1.20)
TMTA—EQS	71%	3.26 (1.40)
TMTB—EQS	71%	2.55 (1.34)
Letter Fluency—EQS	100%	3.21 (1.17)
Category Fluency—EQS	100%	2.42 (1.39)
% APOE ϵ 4 carriers	40%	
—/—		73%
—/+		27%
+ /+		0%
% AB1-42 abnormal	100%	66%
% T-TAU abnormal	100%	47%
% P-TAU abnormal	87%	11%

EQS, Equivalent Score; SD, standard deviation; MMSE, Mini Mental State Examination; AVLT, Rey Auditory Verbal Learning Test; TMT, trail-making test; AB1-42, amyloid beta; T-TAU, total Tau; P-TAU, phosphorylated Tau.

preparation of the multiple single-case analyses on the Piramal data set.

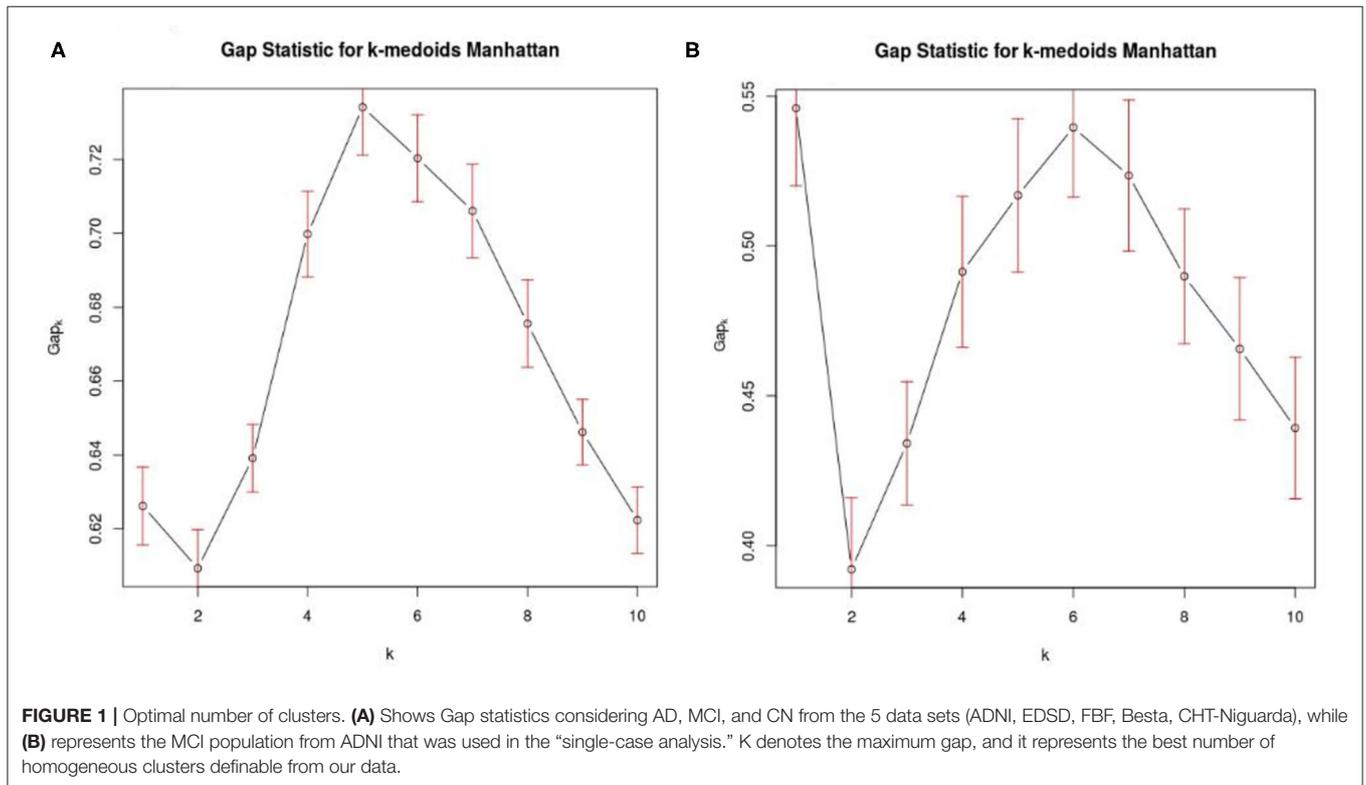
Categorization, Clustering, and Classification Results

For the “group analysis,” starting from the original diagnosis (first category) and the 12 previously defined second category markers (i.e., neurodegeneration appraised by regional atrophy), the Gap statistics identified that five homogeneous classes would be appropriate (**Figure 1A**). The PAM assigned each subject to one of the five clusters. For visualization purposes, the two most important components were used to draw the data points represented in **Figure 2A**. Noticeably, the leftmost cluster (i.e., cluster number 1) was the cluster with the maximum percentage of AD (84% AD, 16% MCI, 0% CN), the lowest MMSE score (21.2 ± 4.0), the highest prevalence of APOE4 carriers (71%), and the highest level of abnormal AB1-42, T-TAU, and P-TAU proteins in the CSF, while the rightmost cluster (i.e., cluster number 3) contained primarily CN (5% AD, 26% MCI, 69% CN) and had the highest MMSE score (27.5 ± 2.4), the lowest prevalence of APOE4 carriers (24%), and the lowest level of abnormal AB1-42, T-TAU, and P-TAU proteins in the CSF. The CCC identified 24 potential informative features, in addition to the 12 features of the second category, that defined the disease fingerprint of the five clusters. A matching matrix summarizing the results of predicting the five classes using just the 24 potential markers is reported in **Figure 3A**. The values on the diagonal (i.e., true positive) were higher than the others, which suggests that these potential features may be used to fairly define the definitive disease fingerprint of each patient. To understand the meaning of the five new created classes, we present a distinctive profiles plot in **Figure 4A**. The median line of each class is well segregated from the others, and the confidence intervals for each biomarker rarely overlapped to the others, which suggests

TABLE 4 | Results from GB classifiers with different validation techniques.

Validation technique	Features used	Train set	Test set	Accuracy	Precision	Recall
(Nested) 10-fold cross-validation	143 (SCD., NPSY., MRI, GEN., CSF)	All	All	85 ± 4%	91 ± 5%	91 ± 5%
	140 (SCD., NPSY., MRI, GEN.)	All	All	69 ± 1%	79 ± 8%	79 ± 7%
Holdout validation	143 (SCD., NPSY., MRI, GEN., CSF)	ADNI + EDSO	FBF + Besta + CHT-Niguarda	94.63%	94.32%	94.70%
	143 (SCD., NPSY., MRI, GEN., CSF)	ADNI + EDSO + Besta + CHT-Niguarda	FBF	95.18%	95.41%	95.41%
	143 (SCD., NPSY., MRI, GEN., CSF)	ADNI + EDSO + FBF + CHT-Niguarda	Besta	91.52%	91.33%	88.77%
	143 (SCD., NPSY., MRI, GEN., CSF)	ADNI + EDSO + FBF + Besta	CHT-Niguarda	95.23%	96.43%	97.22%
(Nested) 5-fold cross-validation	143 (SCD., NPSY., MRI, GEN., CSF)	ADNI (stable & conv)	ADNI (stable & conv)	62 ± 1%	66 ± 1%	70 ± 1%

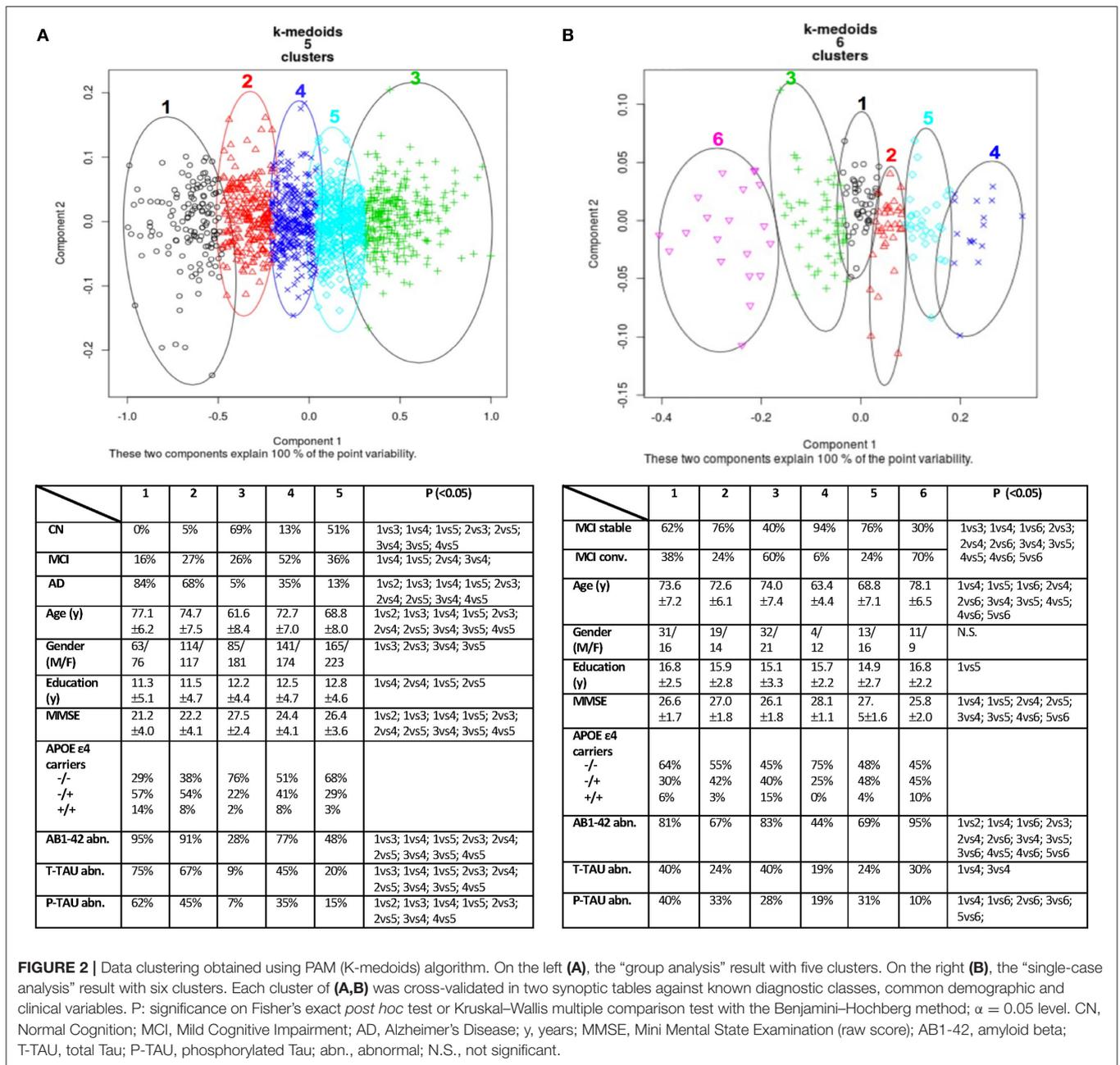
SCD, sociodemographic variables; NPSY, neuropsychological variables; GEN, genetic variables. All, ADNI + EDSO + FBF + Besta + CHT-Niguarda data sets; stable, 118 MCI stable subjects from ADNI; conv, 80 MCI converter to AD or other form of dementia from ADNI.



good cluster separations and new informative purely data-driven diagnostic classes.

For the “single-case analysis,” the Gap statistics identified six subclasses (**Figure 1B**) from the 198 MCI (118 stable and 80 converter) patients acquired from ADNI (**Figure 2B**). The leftmost cluster (i.e., cluster number 6) was the cluster with the maximum percentage of MCI converters (70% converter, 30% stable), the lowest MMSE score (25.8 ± 2.0), the highest

prevalence of both APOE4 carrier (55%), and AB1-42 positive (95%) subjects, while the rightmost cluster (i.e., cluster number 4) contained primarily MCI stable (6% converter, 94% stable) and had the highest MMSE score (28.1 ± 1.1), the lowest prevalence of APOE4 carriers (25%), and the lowest level of abnormal AB1-42, T-TAU, and P-TAU proteins in the CSF. The matching matrix (**Figure 3B**) was built considering 17 potential informative markers. The definitive disease fingerprints of the



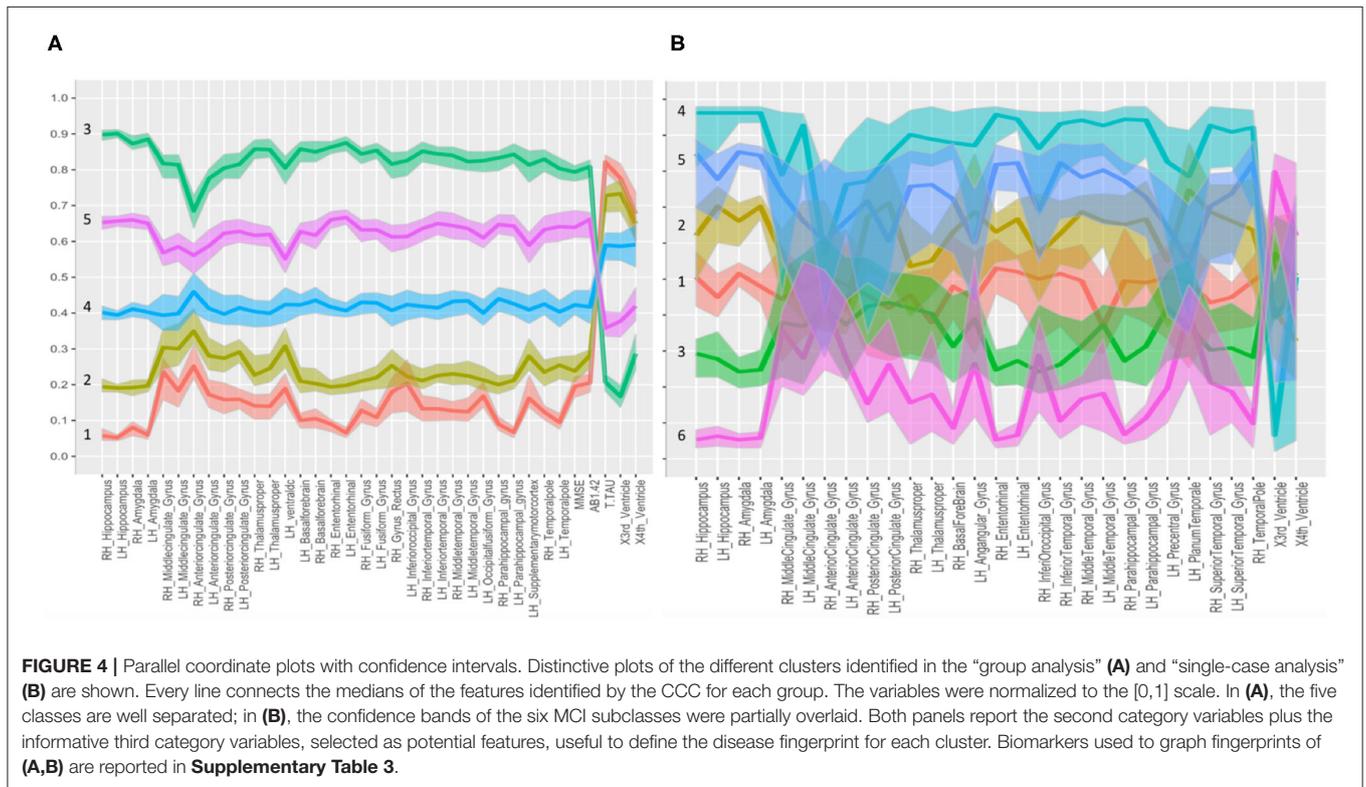
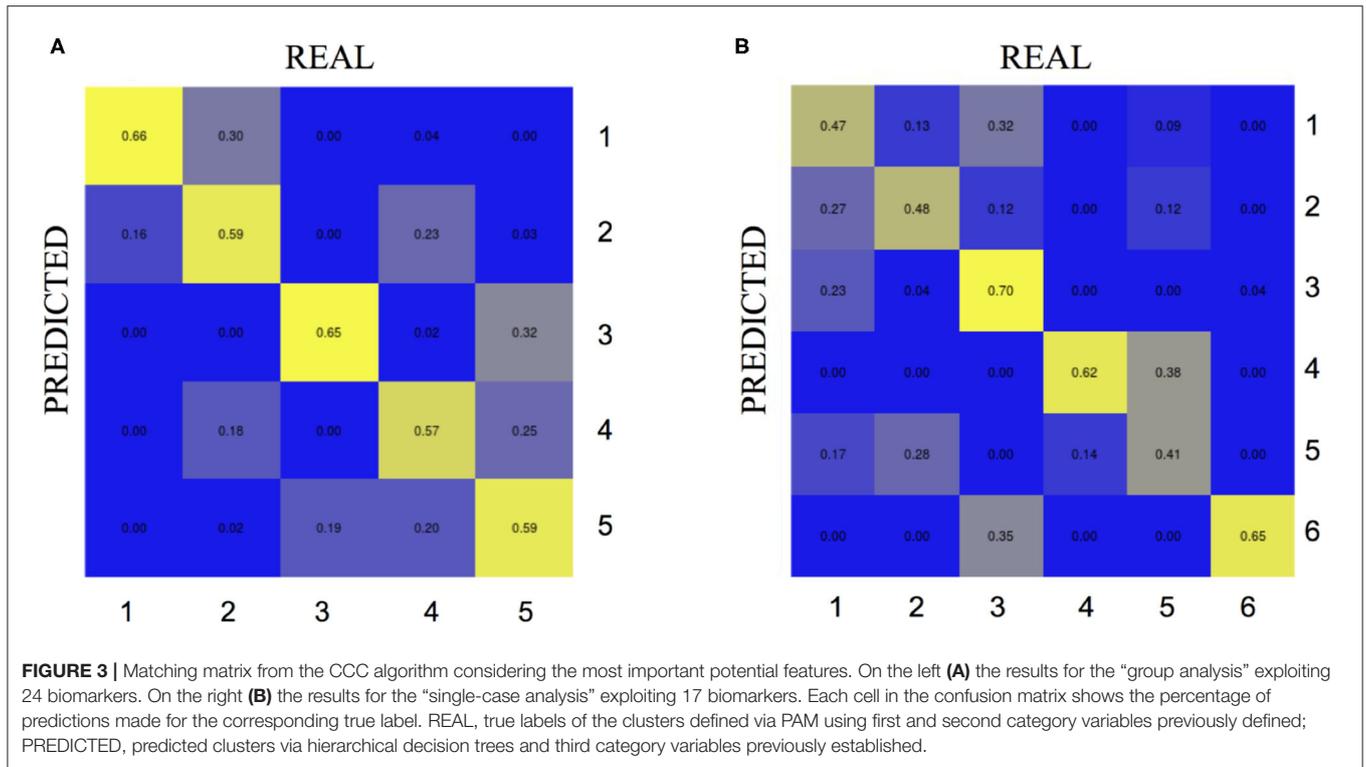
new six classes were graphed combining the 17 markers, with the 12 features of the second category, and using a parallel coordinate plot (Figure 4B).

Clinical Utility of the ML Tools in the MIP

The information provided by the MIP tools did not modify the inter-rater agreement of the four physicians. The inter-rater concordance on the etiologic diagnosis (pathophysiologically AD related vs. not pathophysiologically AD related) was moderate, with a Fleiss' kappa of 0.521 ($p < 0.001$).

The MIP information induced a change in the original etiological hypothesis in three evaluations. We registered

10 changes in the hypothesis of subject conversion with a nonsignificant Fisher's exact test equal to 0.909 at alpha = 0.05 level. The registered increment in diagnostic confidence before and after the MIP information disclosure was equal to +5.99 percentage points (Figure 5A). In Figure 5B, the DC is described in a scatter plot. Points drawn in cooler colors, which denote non-pathophysiologically AD-related subjects, are in most of the cases located above the bisector of the plane, thus showing that the MIP information increased the DC of the physician's prediction. Points drawn in warmer colors (i.e., pathophysiologically AD related) are above the bisector, for lower initial DC values, and become closer to the bisector once the



initial DC value increases. Therefore, the impact of the MIP was more evident when the initial DC of MCI conversion to AD is low (<70%).

When directly asked about the perceived impact of the MIP tools on the final DC, in 20% of cases clinicians reported that the additional information did not change their initial DC. In 36% and in 29% of cases, they felt “somewhat” or “slightly” impacted, respectively. In 15% of cases, they significantly changed their initial belief.

DISCUSSION

The present study demonstrated that the MIP provided dedicated tools for CN, MCI, and AD classification reaching in our “group analysis” similar performance to the state of the art ML algorithms (54–56). We used the supervised GB classifier based on the combination of sociodemographic, MMSE, APOE ϵ 4 genotype, CSF, and MRI data to predict the patient’s stage in the entire disease spectrum with fair accuracy. Such level of accuracy was expected and can be explained because the clinical data used for prediction partially correlated with those used for making the supervised diagnosis at baseline. GB performances were higher for the hospital holdout tasks than in cross-validation and reached a maximum peak of 95% of accuracy in two out of the three clinical cohorts. The following possible reasons explain these results: (i) the evaluation depends on which data points end up in the training set and which end up in the test set, furthermore the evaluation often may be significantly different depending on the hospital cohort subdivision; (ii) the holdout strategy is subjected to higher variance because of the smaller size of data, and the samples used might not be always representative; and (iii) the holdout method might provide only an estimate of the “true error rate” (accuracy) of the classifier.

Considering the power of the MIP to federate features from a limitless number of data sets across the globe, our results are encouraging and exciting. Indeed, the MIP platform will need to be further tested in future studies with other cohorts. To quantitatively measure the performance of the GB in discriminating MCI stable vs. converters, we used the longitudinal ADNI data set, which represented an indirect validation and a plausibility test of the MIP tool granting a 62% of accuracy. This level of performance is consistent with the results proposed in the literature (57–59) using similar features available in our study. It is realistic that GB performances were suboptimal when considering a complex classification task, such as the prediction of MCI conversion, compared to the earlier distinction between well-established diagnostic classes, i.e., CN, MCI, and AD. GB is an advanced and powerful ML tool; however, it cannot provide a real change in the diagnostic paradigm. This issue is naturally ingrained in the supervised algorithm design that cannot allow the definition of new diagnostic classes. However, GB was judged to be informative for the physicians in the “single-case analysis.”

The use of the MIP with the CCC algorithm, based on its agnostic data-mining approach called known-group validity, identified finer clusters and provided a newer data-led

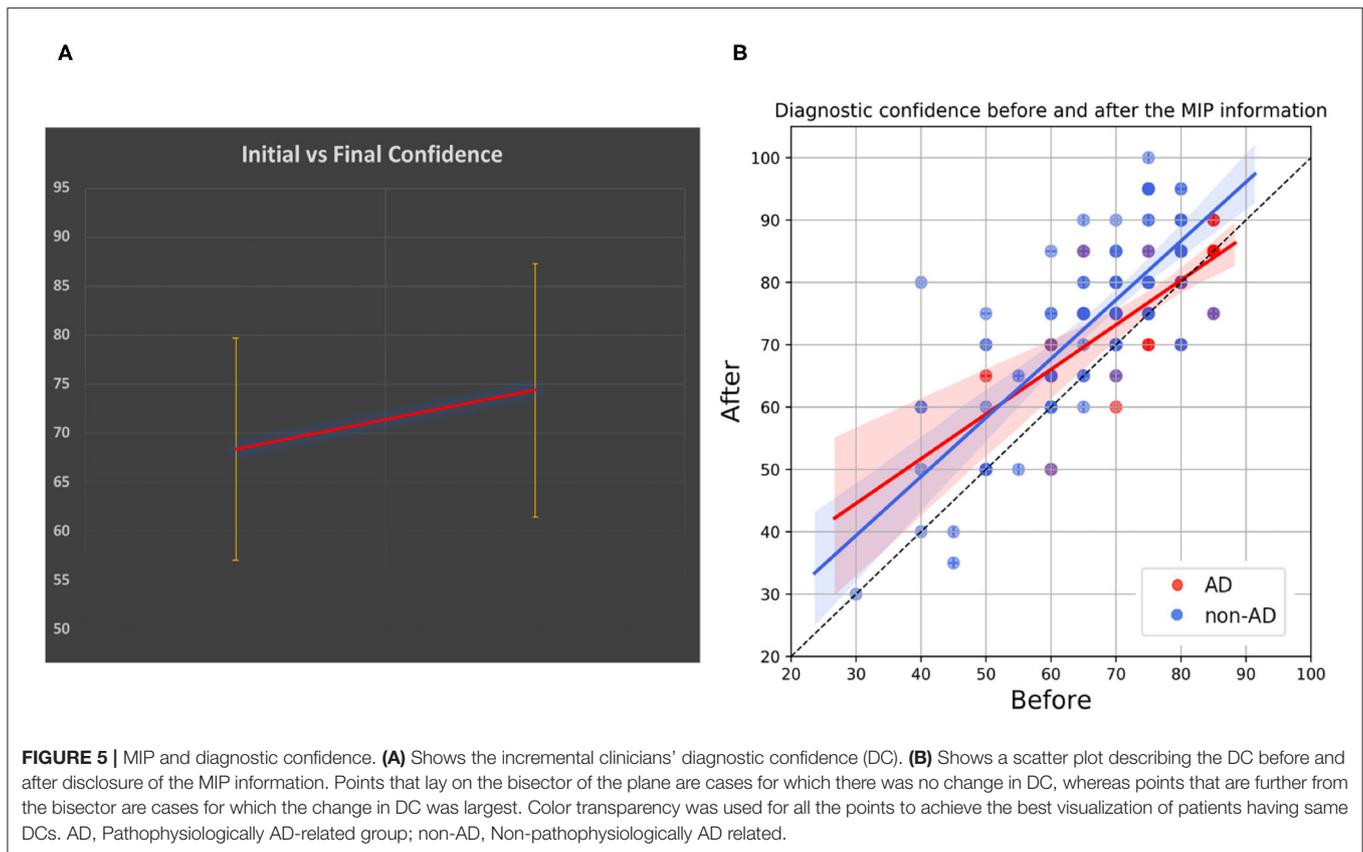
stratification of disease based on each patient’s brain features. We presented a semi-supervised approach that separated patients into homogenous groups according to well-known imaging biomarkers of neurodegeneration. Notably, in the “group analysis,” we found five classes, which were identified in another recent study (60), that might represent a more careful distinction of subjects compared with the three classical broad diagnostic groups, such as CN, MCI, and AD. We enriched the conventional imaging biomarker set of the second category with 24 additional potential markers to demonstrate the consistency of the new classes. Overall, the matching matrix produced interestingly results in which many true positive predictions were located on the diagonal of the table, which gave reliability to the selected potential features of the third category to discriminate the five clusters. The matching matrix showed small prediction errors because percentage values outside the diagonal were generally small. One possible explanation for the mismatched cases is that there were some partially interrelated features among the features used for the classification. CCC defined groups that were characterized by sets of quantifiable biological and clinical variables (alias disease fingerprint) that well defined the biological makeup of the disorders.

The same approach was used in the “single-case analysis” to obtain six classes, with 17 potential features that were useful to define the disease fingerprint of the MCI subgroups in combination with the conventional imaging biomarker set of the second category. We do not claim that our findings present the best current views on the problem. Although the perceived impact of clinicians was in favor of the MIP tools, how well the CCC strategy identified the risk of conversion and the etiology must be yet determined because information about follow-ups and postmortem examinations is not provided in the Piramal database. Future studies must thus be designed to validate this stimulating finding. Moreover, expert knowledge in enlightening the data, such as diversification of questions to different cognitive domains or other biomarker measurements, could help create even more refined clusters of disease arrangement.

In the “single-case” part of this study, MIP provided an added value to clinicians’ DC even when used on top of the traditional diagnostic workup. MIP in combination with neuropsychological assessments and CSF biomarkers had not a trivial impact with an increment equal to +5.99% in the final DC. This result was consolidated considering the perception of the physicians involved in the scoring of MCI subjects who declared that MIP information influenced their DC with an impact that was rated from “slight” to “significant” in 80% of the cases.

Because the MIP platform is by definition a simple, quantitative, reproducible tool that requires fast training, it may provide an important added value in the diagnostic process of the dementia.

Few discordant and contradictory MIP results were registered by physicians between CCC and GB. These were primarily due to algorithm implementation differences in terms of ML approaches and different selection mechanisms of biomarkers to be used. CCC performed a feature pruning mechanism, GB instead was more robust to overfitting and tried to maximize the pathological information using all of the available features. Agreement



between MIP tools is thus subject to technical differences, which may lead to slightly different diagnostic conclusions on the same patient. Therefore, the physician is absolutely essential and not replaceable by any artificial intelligence system, not even the most advanced (61, 62). However, discordant cases may also be due to borderline results.

This said, the MIP could contribute to help physicians in their daily workup and in their patient management and it may be used to test other clinical hypotheses. One of the main advantages of adopting MIP and its integrated ML tools is to combine hundreds of pieces of information (i.e., features) of different nature in a few graphs for the benefit of clinicians, which simplifies how to process the wealth of information available.

Analyzing health and especially hospital data, which are normally much noisier than research data, requires researchers to face some challenges. Some of the major challenges are reported hereinafter. Compensatory mechanisms may obscure the linkage between biological markers and disease (63). For example, two subjects with the same brain images do not necessarily share the same clinical manifestation. It is not only due to the complexity of the disease or the inefficiency of the marker but also by the fact that different compensatory mechanisms exist from one subject to another, which generate a miscellaneous effect on the clinical phenotype. The rate of agreement between physicians is generally modest, which means that supervised ML approaches may be partially invalidated by a poor initial classification. Recent major pathological reviews reported that the diagnosis of AD was no

better than 60–70% accurate (4, 5). This had generated a great debate on the potentially greater usefulness of semi-supervised or completely data-driven approaches. Big-data analyses may increase the possibility to tease out irrelevant biomarkers that were identified by chance in reduced populations. To tackle this methodological problem, a well-founded and validated selection process for features and patients must be performed.

The CCC approach clearly showed the presence of disease classes beyond clinical ones in the overall group of patients and within the MCI group itself. Notably, the overall patient group was divided into classes with clearly defined features that did not overlap with each other and maintained a specific order (i.e., in “group analysis,” group 3 was above all other groups for all features, and group 1 was characterized by the lowest values for all features, with the inversion of the 4th ventricle size). Therefore, the disease fingerprint we can derive for each patient is appealing and may be beneficial for their selection in the future clinical trials, as long as a deeper phenotyping of the identified clusters will be available and confirmed by the clinicians.

Unfortunately, the number of new drugs entering the market is scarce, and many big-pharma companies have recently stopped investing in this area. This situation is a direct consequence of the lack of a clear causative understanding of AD, and it is difficult to find new treatments for brain diseases in this scenario. Therapies in AD normally focus on cholinesterase inhibitors, which are only given after the onset of symptoms.

To change this perspective toward a new paradigm, the MIP platform is gathering a large number of patients and healthy controls to characterize the disease from quantifiable biomarkers instead of gathering the patients' referred symptoms and from few non-standardized instrumental assessments.

We showed how the MIP platform could provide accurate and innovative ways to detect, stratify, and classify patients. We identified possible subject-specific biological signatures that are not captured by the traditional and oversimplified diagnostic categories. These fingerprints should clarify the researchers and physicians on the causes and mechanisms of AD to highlight targets for effective personalized treatment approaches and upcoming prevention studies.

The present study was cross-sectional, and it is still far from a real big analytics experiment despite the gathering of a large number of multidomain features from different cohorts of patients. Future efforts should be made to identify sets of biomarkers that distinguish different disease trajectories longitudinally. This goal requires the integration of data sets with follow-ups within the MIP and appropriate spatiotemporal mathematical models (64–66), which are currently absent in the platform, that may be used as a source of information on biological mechanisms that drive AD progression.

Several unanswered questions must be clarified in the future, such as the following: Is the disease signature a “guarantee of success” for new drug development? Will biological fingerprint characterization predict disease manifestations more accurately? What is the phenotypic difference between close and distant disease signatures?

The present study has a few limitations. The RIN network includes participating Italian clinical centers, which may represent a selected group that is more likely to make use of innovative biomarkers for diagnosis. In principle, we cannot exclude that different ML tools, hosted within the MIP but not chosen for our specific study, could have different impacts, better or worse, on the classification performances of the group analysis and on the DC we performed. Our study chose GB and CCC algorithms for their implementation readiness inside the MIP platform. However, other promising or most common classifiers such as Random Forest (67), Support Vector Machine, or semi-supervised heuristic approaches (68) should be tested further on. The development of new feature selection strategies to identify relevant and nonredundant feature subsets, innovative ensemble, and deep learning algorithms for clinical classification (69) should be explored too. Except for the MMSE score, in the “single-case analysis” part, aggregated outcomes coming from neuropsychological tests were acquired and dichotomized into normal/abnormal according to different local clinical practices and protocols. The same caveat applies to CSF biomarkers. We cannot claim that different tests and protocols may have different impacts on DC. Furthermore, we included only one biomarker of brain amyloidosis (i.e., CSF Aβ1-42) without considering the amyloid-PET imaging data. We chose not to include amyloid-PET because this examination could not be processed by the MIP because of the lack of an *ad hoc* automatic pipeline of imaging analysis. Indeed, it will be important to integrate this type of pipeline so that future studies will also take advantage

of multimodal MRI and PET imaging information. Lastly, the limited sample size in the single-case analysis of MCI patients warrants replication in larger studies.

CONCLUSION

Thanks to the MIP platform, we are assisting to a radical change in neuroscience that is symbolized by moving away from traditional syndromic diagnosis toward diagnoses based on biological signatures, while not ruling out the importance of clinician supervision. Advanced statistics, notably data mining and ML tools, are the armamentarium for this paradigm shift. The novel approach we demonstrated does not constitute a substitute for the classical hypothesis-led approach conducted in neuroscience so far. Rather, it is a complementary methodology that allows a better understanding of the complexity of the brain and its diseases. The present study also demonstrated an effect on DC and an influence on the physicians' clinical thinking and decision making when the MIP outcomes were added to other biomarkers.

The disease signature, emerging from big-data analyses and innovative ML approaches, seems to play a promising role in patients' stratification within future clinical trials or observational studies. This methodology may help to identify new targets for intervention, guide better care, and lead to precision diagnostics.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.neugrid2.eu/index.php/data-portfolio/>; <http://adni.loni.usc.edu/data-samples/access-data/>; <https://services.humanbrainproject.eu/oidc/login>.

ETHICS STATEMENT

The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AR: conceptualization, methodology, software, formal analysis, investigation, writing, and visualization. SD: methodology, software, formal analysis, investigation, resources, data curation, and writing. FP: software and formal analysis. SG: methodology and validation. CM: resources and data curation. GC and GS: software. PT and GB: validation. AN: software and resources. MB: resources. MR: methodology. SF: project administration. CG: writing. FT: supervision. GF: resources and supervision. PR: software and supervision. J-FD: writing and supervision. FK: software, data curation, and supervision. SC: methodology, validation, writing, and supervision. ED'A: conceptualization, software, investigation, resources, writing - review - editing, and supervision. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

Data used in preparation of this article were partially obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in data analysis or writing of this report. A complete listing of ADNI investigators may be found at https://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf. ADNI data were funded by the Alzheimer's Disease Neuroimaging Initiative (National Institutes of Health grant U01 AG024904) and Department of Defense Alzheimer's Disease Neuroimaging Initiative (Department of Defense award W81XWH-12-2-0012). The Alzheimer's Disease Neuroimaging Initiative was funded by the National Institute on Aging and the National Institute of Biomedical Imaging and Bioengineering, and through contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd. and its affiliated company Genentech Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research and Development LLC; Johnson & Johnson Pharmaceutical Research & Development LLC; Lumosity; Lundbeck; Merck and Co Inc.; Meso Scale Diagnostics LLC; NeuroRx Research; Neuro-track Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. Private sector contributions are facilitated by the Foundation for the National Institutes of Health. The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's

Therapeutic Research Institute at the University of Southern California. Alzheimer's Disease Neuroimaging Initiative data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. ESDS data used in the preparation of this article were originally obtained from Prof. Stefan Teipel (DZNE) and stored in the NeuGRID platform (<https://www.neugrid2.eu>) funded by grant 283562 from the European Commission and then shared with the HBP MIP initiative. The present research has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under the specific grant agreement 785907 (Human Brain Project Rump-UP, SGA1, SGA2). The research leading to these results has also received funding from the European Community's Horizon 2020 research and innovation programme under grant agreement no 871643: Morphemic. The present study was supported by Piramal Imaging study (grant number 115952) and by the Italian Network of Neuroscience and Neurorehabilitation (RIN) neuroimaging initiatives. This study was funded by a National funding by Italian Ministry of Health in the framework of the grant RETE IRCCS DI NEUROSCIENZE E NEURORIABILITAZIONE (Imaging Project - RRC-2016-2361095; RRC-2017-2364915; RRC-2018-2365796). Special thanks to Tel Aviv University in the person of Alexis Mitelpunkt, Tal Galili, Mira Marcus-Kalish, and Yoav Benjamini for the provision of the CCC algorithm in the MIP platform.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2020.01021/full#supplementary-material>

REFERENCES

- Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol.* (2007) 6:734–46. doi: 10.1016/S1474-4422(07)70178-3
- Dubois B, Feldman HH, Jacova C, Cummings JL, Dekosky ST, Barberger-Gateau P, et al. Revising the definition of Alzheimer's disease: a new lexicon. *Lancet Neurol.* (2010) 9:1118–27. doi: 10.1016/S1474-4422(10)70223-4
- Beach TG, Monsell SE, Phillips LE, Kukull W. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010. *J Neuropathol Exp Neurol.* (2012) 71:266–73. doi: 10.1097/NEN.0b013e31824b211b
- Nelson PT, Alafuzoff I, Bigio EH, Bouras C, Braak H, Cairns NJ, et al. Correlation of Alzheimer disease neuropathologic changes with cognitive status: a review of the literature. *J Neuropathol Exp Neurol.* (2012) 71:362–81. doi: 10.1097/NEN.0b013e31825018f7
- Jack CR Jr, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, et al. NIA-AA research framework: toward a biological definition of Alzheimer's disease. *Alzheimers Dement.* (2018) 14:535–62. doi: 10.1016/j.jalz.2018.02.018
- Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. 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. *Alzheimers Dement.* (2011) 7:270–9. doi: 10.1016/j.jalz.2011.03.008
- Frackowiak R, Markram H. The future of human cerebral cartography: a novel approach. *Philos Trans R Soc Lond B Biol Sci.* (2015) 370:20140171. doi: 10.1098/rstb.2014.0171
- Redolfi A, Bosco P, Manset D, Frisoni GB, neuGRID consortium. Brain investigation and brain conceptualization. *Funct Neurol.* (2013) 28:175–90. doi: 10.11138/FNeur/2013.28.3.175
- Cui J, Zufferey V, Kherif F. *In-vivo* brain neuroimaging provides a gateway for integrating biological and clinical biomarkers of Alzheimer's disease. *Curr Opin Neurol.* (2015) 28:351–7. doi: 10.1097/WCO.0000000000000225
- Frackowiak R, Ailamaki A, Kherif F. Federating and integrating what we know about the brain at all scales: computer science meets the clinical neurosciences. In: Buzsáki G, Christen Y, editors. *Micro-, Meso- and Macro-Dynamics of the Brain*. Cham, CH: Springer (2016). p. 157–70.
- Aarestrup FM, Albeyatti A, Armitage WJ, Auffray C, Augello L, Balling R, et al. Towards a European health research and innovation cloud (HRIC). *Genome Med.* (2020) 12:18. doi: 10.1186/s13073-020-0713-z
- Frisoni GB, Redolfi A, Manset D, Rousseau MÉ, Toga A, Evans AC. Virtual imaging laboratories for marker discovery in neurodegenerative diseases. *Nat Rev Neurol.* (2011) 7:429–38. doi: 10.1038/nrneuro.2011.99
- Laboratory of Neuro Imaging. *USC (LONI)*. (2009). Available online at: <https://loni.usc.edu/>
- neuGRID*. (2011). Available online at: <https://neugrid2.eu/>
- Redolfi A, McClatchey R, Anjum A, Zijdenbos A, Manset D, Barkhof F, et al. Grid infrastructures for computational neuroscience: the neuGRID example. *Fut Neurol.* (2009) 4:703–22. doi: 10.2217/fnl.09.53

16. Redolfi A, Manset D, Barkhof F, Wahlund LO, Glatard T, Mangin JF, et al. Head-to-head comparison of two popular cortical thickness extraction algorithms: a cross-sectional and longitudinal study. *PLoS ONE*. (2015) 10:e0117692. doi: 10.1371/journal.pone.0117692
17. *CBRAIN*. (2008). Available online at: <http://cbrain.mcgill.ca/>
18. Oliveira JL, Trifan A, Bastião Silva LA. EMIF Catalogue: a collaborative platform for sharing and reusing biomedical data. *Int J Med Inform*. (2019) 126:35–45. doi: 10.1016/j.ijmedinf.2019.02.006
19. Ten Kate M, Redolfi A, Peira E, Bos I, Vos SJ, Vandenberghe R, et al. MRI predictors of amyloid pathology: results from the EMIF-AD Multimodal Biomarker Discovery study. *Alzheimers Res Ther*. (2018) 10:100. doi: 10.1186/s13195-018-0428-1
20. Salles A, Bjaalie JG, Evers K, Farisco M, Fothergill BT, Guerrero M, et al. The human brain project: responsible brain research for the benefit of society. *Neuron*. (2019) 101:380–4. doi: 10.1016/j.neuron.2019.01.005
21. Auffray C, Balling R, Barroso I, Bencze L, Benson M, Bergeron J, et al. Making sense of big data in health research: Towards an EU action plan [published correction appears in *Genome Med*. (2016) 8(1):118]. *Genome Med*. (2016) 8:71. doi: 10.1186/s13073-016-0376-y
22. Draganski B, Kherif F, Damian D, Demonet JF, MemoNet consortium. A nation-wide initiative for brain imaging and clinical phenotype data federation in Swiss university memory centers. *Curr Opin Neurol*. (2019) 32:557–63. doi: 10.1097/WCO.0000000000000721
23. *RIN*. (2017). Available online at: <https://www.reteneuroscienze.it/en/>
24. Sivarajah U, Kamal MM, Irani Z, Weerakkody V. Critical analysis of big data challenges and analytical methods. *J Business Res*. (2017) 70:263–86. doi: 10.1016/j.jbusres.2016.08.001
25. Visser PJ, Verhey FR, Boada M, Bullock R, De Deyn PP, Frisoni GB, et al. Development of screening guidelines and clinical criteria for predementia Alzheimer's disease. The DESCRIPA Study. *Neuroepidemiology*. (2008) 30:254–65. doi: 10.1159/000135644
26. Zuo XN, Xing XX. Effects of non-local diffusion on structural MRI preprocessing and default network mapping: statistical comparisons with isotropic/anisotropic diffusion. *PLoS ONE*. (2011) 6:e26703. doi: 10.1371/journal.pone.0026703
27. Dalca AV, Bouman KL, Freeman WT, Rost NS, Sabuncu MR, Golland P. Population based image imputation. *Inf Process Med Imaging*. (2017) 10265:659–71. doi: 10.1007/978-3-319-59050-9_52
28. Mattsson N, Zetterberg H, Hansson O, Andreasen N, Parnetti L, Jonsson M, et al. CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. *JAMA*. (2009) 302:385–93. doi: 10.1001/jama.2009.1064
29. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer's disease. *Neurology*. (1984) 34:939–44. doi: 10.1212/WNL.34.7.939
30. Morris JC, Heyman A, Mohs RC, Hughes JP, van Belle G, Fillenbaum G, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*. (1989) 39:1159–65. doi: 10.1212/WNL.39.9.1159
31. Dubois B, Feldman HH, Jacova C, Hampel H, Molinuevo JL, Blennow K, et al. Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria [published correction appears in *Lancet Neurol*. (2014) 13(8):757]. *Lancet Neurol*. (2014) 13:614–29. doi: 10.1016/S1474-4422(14)70090-0
32. *CCC*. Available online at: <https://github.com/HBPMedical/CCC>
33. Friedman JH. Stochastic gradient boosting. *Comput Stat. Data Anal*. (2002) 38:367–78. doi: 10.1016/S0167-9473(01)00065-2
34. Mitelpunkt A, Galili T, Shachar N, Marcus-Kalish M, Benjamini Y. Categorize, cluster & classify - the 3C strategy applied to Alzheimer's disease as a case study. *Healthinf*. (2015) 1:566–73. doi: 10.5220/0005275705660573
35. *HBP Development Repository*. Available online at: <https://hbpmedical.github.io/deployment-pack/>
36. Bruggen K, Grothe MJ, Dyrba M, Fellgiebel A, Fischer F, Filippi M, et al. The European DTI study on dementia - a multicenter DTI and MRI study on Alzheimer's disease and mild cognitive impairment. *Neuroimage*. (2017) 144(Pt B):305–8. doi: 10.1016/j.neuroimage.2016.03.067
37. Petersen RC, Aisen PS, Beckett LA, Donohue MC, Gamst AC, Harvey DJ, et al. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. *Neurology*. (2010) 74:201–9. doi: 10.1212/WNL.0b013e3181cb3e25
38. Ramusino MC, Altomare D, Assal F, Mendes A, Costa A, Morbelli S, et al. Frisoni, The incremental value of amyloid pet versus Csf biomarkers for the diagnosis of Alzheimer's Disease (India-Fbb Study). *Alzheimer's & Dement*. (2018) 14 (Suppl.):P17–8. doi: 10.1016/j.jalz.2018.06.2070
39. *Neuromorphometric*. Available online at: <http://www.neuromorphometrics.com/>
40. *Neuromorphometric*. Available online at: <https://www.nitrc.org/projects/manuallabels/>
41. Reite M, Reite E, Collins D, Teale P, Rojas DC, Sandberg E. Brain size and brain/intracranial volume ratio in major mental illness. *BMC Psychiatry*. (2010) 10:79. doi: 10.1186/1471-244X-10-79
42. Kang JH, Vanderstichele H, Trojanowski JQ, Shaw LM. Simultaneous analysis of cerebrospinal fluid biomarkers using microsphere-based xMAP multiplex technology for early detection of Alzheimer's disease. *Methods*. (2012) 56:484–93. doi: 10.1016/j.ymeth.2012.03.023
43. Sjögren M, Vanderstichele H, Agren H, Zachrisson O, Edsbacke M, Wikkelso C, et al. Tau and Abeta42 in cerebrospinal fluid from healthy adults 21–93 years of age: establishment of reference values. *Clin Chem*. (2001) 47:1776–81. doi: 10.1093/clinchem/47.10.1776
44. Frisoni GB, Prestia A, Zanetti O, Galluzzi S, Romano M, Cotelli M, et al. Markers of Alzheimer's disease in a population attending a memory clinic. *Alzheimers Dement*. (2009) 5:307–17. doi: 10.1016/j.jalz.2009.04.1235
45. Falahati F, Ferreira D, Soiminen H, Mecocci P, Vellas B, Tsolaki M, et al. The effect of age correction on multivariate classification in Alzheimer's Disease, with a focus on the characteristics of incorrectly and correctly classified subjects. *Brain Topogr*. (2016) 29:296–307. doi: 10.1007/s10548-015-0455-1
46. Xu Z, Huang G, Weinberger KQ, Zheng AX. Gradient boosted feature selection. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY (2014).
47. Khanna S, Domingo-Fernández D, Iyappan A, Emon MA, Hofmann-Apitius M. Using multi-scale genetic, neuroimaging and clinical data for predicting Alzheimer's Disease and reconstruction of relevant biological mechanisms. *Sci Rep*. (2018) 8:11173. doi: 10.1038/s41598-018-29433-3
48. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. (2011) 12:2825–30.
49. Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. *J R Stat Soc B*. (2001) 63:411–23. doi: 10.1111/1467-9868.00293
50. Park HS, Jun CH. A simple and fast algorithm for K-medoids clustering. *Exp Syst Appl*. (2009) 36:3336–41. doi: 10.1016/j.eswa.2008.01.039
51. Cutler A, Cutler DR, Stevens JR. Random forests. In: Zhang C and Ma Y, editors. *Ensemble Machine Learning: Methods and Applications*. Boston: Springer US (2012). p. 157–75.
52. Breiman, L. Random forests. *Mach Learn*. (2001) 45:5–32. doi: 10.1023/A:1010933404324
53. Capitani E, Laiacina M. Composite neuropsychological batteries and demographic correction: standardization based on equivalent scores, with a review of published data. The Italian Group for the Neuropsychological Study of Ageing. *J Clin Exp Neuropsychol*. (1997) 19:795–809. doi: 10.1080/01688639708403761
54. Riedel BC, Daianu M, Ver Steeg G, Mezher A, Salminen LE, Galstyan A, et al. Uncovering biologically coherent peripheral signatures of health and risk for Alzheimer's Disease in the aging brain. *Front Aging Neurosci*. (2018) 10:390. doi: 10.3389/fnagi.2018.00390
55. Aghili M, Tabarestani S, Adjouadi M, Adeli E. Predictive modeling of longitudinal data for Alzheimer's Disease Diagnosis Using RNNs. In: *First International Workshop, PRIME 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings* (2018).
56. Salvatore C, Battista P, Castiglioni I. Frontiers for the early diagnosis of AD by means of MRI brain imaging and support vector machines. *Curr Alzheimer Res*. (2016) 13:509–33. doi: 10.2174/1567205013666151116141705
57. Ramirez J, Górriz JM, Ortiz A, Martínez-Murcia FJ, Segovia F, Salas-Gonzalez D, et al. Ensemble of random forests One vs. Rest classifiers for

- MCI and AD prediction using ANOVA cortical and subcortical feature selection and partial least squares. *J Neurosci Methods*. (2018) 302:47–57. doi: 10.1016/j.jneumeth.2017.12.005
58. Adaszewski S, Dukart J, Kherif F, Frackowiak R, Draganski B. How early can we predict Alzheimer's disease using computational anatomy? *Neurobiol Aging*. (2013) 34:2815–26. doi: 10.1016/j.neurobiolaging.2013.06.015
 59. Salvatore C, Cerasa A, Battista P, Gilardi MC, Quattrone A, Castiglioni I. Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach. *Front Neurosci*. (2015) 9:307. doi: 10.3389/fnins.2015.00307
 60. Toschi N, Lista S, Baldacci F, Cavedo E, Zetterberg H, Blennow K, et al. Biomarker-guided clustering of Alzheimer's disease clinical syndromes. *Neurobiol Aging*. (2019) 83:42–53. doi: 10.1016/j.neurobiolaging.2019.08.032
 61. Patel UK, Anwar A, Saleem S, Malik P, Rasul B, Patel K, et al. Artificial intelligence as an emerging technology in the current care of neurological disorders. *J Neurol*. (2019) 363:k4563. doi: 10.1007/s00415-019-09518-3
 62. Goldhahn J, Rampton V, Spinaz GA. Could artificial intelligence make doctors obsolete? *BMJ*. (2018) 363:k4563. doi: 10.1136/bmj.k4563
 63. Venetis T, Ailamaki A, Heinis T, Karpathiotakis M, Kherif F, Mitelpunkt A, et al. Towards the identification of disease signatures. In: *International Conference on Brain Informatics and Health. Brain Informatics and Health - BIH. Lecture Notes in Computer Science*, Vol. 9250. Cham: Springer (2015).
 64. Archetti D, Ingala S, Venkatraghavan V, Wottschel V, Young AL, Bellio M, et al. Multi-study validation of data-driven disease progression models to characterize evolution of biomarkers in Alzheimer's disease. *Neuroimage Clin*. (2019) 24:101954. doi: 10.1016/j.nicl.2019.101954
 65. Huizinga W, Poot DHJ, Vernooij MW, Roshchupkin GV, Bron EE, Ikram MA, et al. A spatio-temporal reference model of the aging brain. *Neuroimage*. (2018) 169:11–22. doi: 10.1016/j.neuroimage.2017.10.040
 66. Koval I, Schiratti JB, Routier A, Bacci M, Colliot O, Allasonnière S, et al. Spatiotemporal propagation of the cortical atrophy: population and individual patterns. *Front Neurol*. (2018) 9:235. doi: 10.3389/fneur.2018.00235
 67. Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's Disease: a systematic review. *Front Aging Neurosci*. (2017) 9:329. doi: 10.3389/fnagi.2017.00329
 68. Khajehnejad M, Saatlou FH, Mohammadzade H. Alzheimer's Disease early diagnosis using manifold-based semi-supervised learning. *Brain Sci*. (2017) 7:109. doi: 10.3390/brainsci7080109
 69. Sarica A, Cerasa A, Quattrone A, Calhoun V. Editorial on special issue: machine learning on MCI. *J Neurosci Methods*. (2018) 302:1–2. doi: 10.1016/j.jneumeth.2018.03.011

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Redolfi, De Francesco, Palesi, Galluzzi, Muscio, Castellazzi, Tiraboschi, Savini, Nigri, Bottini, Bruzzone, Ramusino, Ferraro, Gandini Wheeler-Kingshott, Tagliavini, Frisoni, Ryvlin, Demonet, Kherif, Cappa and D'Angelo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.