



# A Machine Learning Approach for the Differential Diagnosis of Alzheimer and Vascular Dementia Fed by MRI Selected Features

Gloria Castellazzi<sup>1,2,3\*</sup>, Maria Giovanna Cuzzoni<sup>4</sup>, Matteo Cotta Ramusino<sup>5,6</sup>, Daniele Martinelli<sup>6,7</sup>, Federica Denaro<sup>4</sup>, Antonio Ricciardi<sup>1,8</sup>, Paolo Vitali<sup>9,9</sup>, Nicoletta Anzalone<sup>10</sup>, Sara Bernini<sup>5</sup>, Fulvia Palesi<sup>6</sup>, Elena Sinforiani<sup>5</sup>, Alfredo Costa<sup>5,6</sup>, Giuseppe Miceli<sup>11</sup>, Egidio D'Angelo<sup>6,12</sup>, Giovanni Magenes<sup>3†</sup> and Claudia A. M. Gandini Wheeler-Kingshott<sup>1,3,6\*†</sup>

<sup>1</sup> NMR Research Unit, Department of Neuroinflammation, Faculty of Brain Sciences, Queen Square MS Centre, UCL Queen Square Institute of Neurology, London, United Kingdom, <sup>2</sup> Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy, <sup>3</sup> Brain MRI 3T Research Center, IRCCS Mondino Foundation, Pavia, Italy, <sup>4</sup> Stroke Unit, IRCCS Mondino Foundation, Pavia, Italy, <sup>5</sup> Laboratory of Neuropsychology and Unit of Behavioral Neurology, IRCCS Mondino Foundation, Pavia, Italy, <sup>6</sup> Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy, <sup>7</sup> Headache Center, IRCCS Mondino Foundation, Pavia, Italy, <sup>8</sup> Department of Medical Physics and Biomedical Engineering, Centre for Medical Image Computing, University College London, London, United Kingdom, <sup>9</sup> Radiology Unit, IRCCS Policlinico San Donato, Milan, Italy, <sup>10</sup> Scientific Institute H.S. Raffaele Vita e Salute University, Milan, Italy, <sup>11</sup> Department of Emergency Neurology, IRCCS Mondino Foundation, Pavia, Italy, <sup>12</sup> Brain Connectivity Center, IRCCS Mondino Foundation, Pavia, Italy

## OPEN ACCESS

### Edited by:

Ludovico Minati,  
Tokyo Institute of Technology, Japan

### Reviewed by:

Frithjof Kruggel,  
University of California, Irvine,  
United States  
Maja Puchades,  
University of Oslo, Norway

### \*Correspondence:

Gloria Castellazzi  
gloria.castellazzi@unipv.it  
Claudia A. M. Gandini  
Wheeler-Kingshott  
c.wheeler-kingshott@ucl.ac.uk

†These authors share last authorship

**Received:** 09 January 2020

**Accepted:** 24 April 2020

**Published:** 11 June 2020

### Citation:

Castellazzi G, Cuzzoni MG, Cotta Ramusino M, Martinelli D, Denaro F, Ricciardi A, Vitali P, Anzalone N, Bernini S, Palesi F, Sinforiani E, Costa A, Miceli G, D'Angelo E, Magenes G and Gandini Wheeler-Kingshott CAM (2020) A Machine Learning Approach for the Differential Diagnosis of Alzheimer and Vascular Dementia Fed by MRI Selected Features. *Front. Neuroinform.* 14:25. doi: 10.3389/fninf.2020.00025

Among dementia-like diseases, Alzheimer disease (AD) and vascular dementia (VD) are two of the most frequent. AD and VD may share multiple neurological symptoms that may lead to controversial diagnoses when using conventional clinical and MRI criteria. Therefore, other approaches are needed to overcome this issue. Machine learning (ML) combined with magnetic resonance imaging (MRI) has been shown to improve the diagnostic accuracy of several neurodegenerative diseases, including dementia. To this end, in this study, we investigated, first, whether different kinds of ML algorithms, combined with advanced MRI features, could be supportive in classifying VD from AD and, second, whether the developed approach might help in predicting the prevalent disease in subjects with an unclear profile of AD or VD. Three ML categories of algorithms were tested: artificial neural network (ANN), support vector machine (SVM), and adaptive neuro-fuzzy inference system (ANFIS). Multiple regional metrics from resting-state fMRI (rs-fMRI) and diffusion tensor imaging (DTI) of 60 subjects (33 AD, 27 VD) were used as input features to train the algorithms and find the best feature pattern to classify VD from AD. We then used the identified VD–AD discriminant feature pattern as input for the most performant ML algorithm to predict the disease prevalence in 15 dementia patients with a “mixed VD–AD dementia” (MXD) clinical profile using their baseline MRI data. ML predictions were compared with the diagnosis evidence from a 3-year clinical follow-up. ANFIS emerged as the most efficient algorithm in discriminating AD from VD, reaching a classification accuracy greater than 84% using a small feature pattern. Moreover, ANFIS showed improved classification accuracy when trained with a multimodal input feature data set (e.g., DTI + rs-fMRI metrics) rather than a unimodal feature data set.

When applying the best discriminant pattern to the MXD group, ANFIS achieved a correct prediction rate of 77.33%. Overall, results showed that our approach has a high discriminant power to classify AD and VD profiles. Moreover, the same approach also showed potential in predicting earlier the prevalent underlying disease in dementia patients whose clinical profile is uncertain between AD and VD, therefore suggesting its usefulness in supporting physicians' diagnostic evaluations.

**Keywords:** Alzheimer disease, vascular dementia, machine learning, resting state fMRI, DTI

## INTRODUCTION

Alzheimer disease (AD) is the primary and most frequently diagnosed dementia disease in elderly subjects. At a physiological level, AD is a progressive neurodegenerative disease characterized by the accumulation of amyloid- $\beta$  plaques and tau-related neurofibrillary tangles mainly affecting the prefrontal and mesial-temporal areas of the brain. AD is associated with memory dysfunction and severe cognitive decline caused by a dramatic shrinking of the brain tissues (i.e., atrophy) and neural circuitries (Reitz et al., 2011; Serrano-Pozo et al., 2011). The accurate diagnosis of AD is crucial for patients' management and treatment, but it is often challenging, in particular when AD-like symptoms overlap with cerebrovascular changes, which are also a characteristic trait of vascular dementia (VD) (Groves et al., 2000). From an epidemiological point of view, VD is considered the second most prevalent type of dementia after AD. VD represents a clinical syndrome that includes a wide spectrum of cognitive dysfunctions resulting from brain tissue damage caused by vascular disease that can lead to large artery strokes, small vessel disease (SVD), and other less-frequent vascular lesions (Micieli, 2006; Vinters et al., 2018). From a clinical point of view, VD also represents a great challenge because of its relatively high prevalence and lack of effective treatment options (Baskys and Hou, 2007). Indeed, although cognitive impairment following stroke generally tends to recede, vascular dementia due to SVD is often progressive and may be confused with AD, possibly leading to delays and errors in identifying the best treatment for each individual.

A relevant help in characterizing dementia has come from advanced magnetic resonance imaging (MRI) techniques, such as diffusion tensor imaging (DTI) and resting-state functional magnetic resonance imaging (rs-fMRI) (Agosta et al., 2017; Filippi et al., 2019). A recent DTI study has shown that AD and VD are characterized by distinct patterns of white matter (WM) changes, therefore suggesting DTI parameters as valid biomarkers to investigate the pathogenesis of dementia (Palesi et al., 2018). Several studies have instead used rs-fMRI to investigate the brain functional connectivity (FC) changes caused by neurodegeneration, providing important insights into the pathophysiology of dementia (Castellazzi et al., 2014; Buckley et al., 2017) as well as into the mechanism of disease progression (Dillen et al., 2017). However, despite the large number of MRI studies focused on dementia, the identification of MRI biomarkers to clearly differentiate the AD profile from VD remains difficult.

Improvements in imaging and the advent of affordable powerful computational resources have created a fertile ground for the development of machine learning (ML) approaches, which represent a pool of qualified methods for exploring data to discover already present unknown patterns (Bishop, 2006). Indeed, ML techniques, combined with MRI-derived indices, i.e., quantitative MRI (qMRI), have been used to identify AD subjects from normal elderly people (Long et al., 2017). Other studies have shown that the combination of ML with qMRI represents a suitable approach not only to automatically identify dementia diseases, but also to predict the disease progression as well as the conversion from a mild cognitive impairment (MCI) to a more severe condition, such as AD (Dyrba et al., 2015; Dallora et al., 2017; Pellegrini et al., 2018). Moreover, a recent study showed that ML combined with volumetric measurements derived from structural MRI represents a useful approach for the differential diagnosis between AD and VD (Zheng et al., 2019).

Compared to earlier pieces of work, this study aims to establish the potential of ML algorithms combined with advanced qMRI metrics to automatically discriminate AD from VD. Moreover, we evaluate which algorithms are more suitable in enhancing classification accuracy when using multimodal MR features rather than unimodal data. Finally, we test whether this approach is able to give an earlier and more precise indication (compared to conventional clinical evaluations) about the prevalent underlying disease (i.e., AD rather than VD) in a pool of patients diagnosed with a "mixed" VD-AD dementia (MXD) profile.

## MATERIALS AND METHODS

### Subjects

MRI acquisitions were performed on a total data set of 77 subjects with dementia. Thirty-three subjects diagnosed with AD (age  $72.8 \pm 7.3$ ), and 27 subjects diagnosed with VD (age  $76.6 \pm 7.7$ ) were recruited for the study. A third group of 15 subjects diagnosed with mixed AD-VD dementia (MXD, age  $76.3 \pm 6.7$ ) was included to test the potential of the proposed ML approach in predicting the prevalent underlying dementia disease. AD, VD, and MXD patients were selected on the basis of the NINCDS2-ARDA criteria (McKhann et al., 2011) among those regularly attending the Neurological Institute IRCCS Mondino Foundation (Pavia, Italy). Exclusion criteria were age  $>80$  years and significant medical or neurological (other than AD or VD) or psychiatric disease. Patients with significant

central nervous system (CNS) disorders (e.g., Parkinson's disease and other extra-pyramidal disorders, multiple sclerosis, epilepsy, clinical evidence of acute ischemic or hemorrhagic stroke, known intracranial lesions, systemic causes of subacute cognitive impairment, and/or previous head injury with loss of consciousness) were excluded, too. All subjects were scanned under an institutional review board (IRB) approved protocol after obtaining written informed consent from them or their lawful caregiver.

## Clinical and Neuropsychological Assessment

All subjects underwent clinical and neuropsychological evaluation to assess their global cognitive status using the Mini-Mental State Examination (MMSE) (Folstein et al., 1975) and the following cognitive domains: attention (Stroop test, trail making test A and B, attentive matrices), memory (digit and verbal span, Corsi block-tapping test, logical memory, Rey-Osterrieth complex figure recall, Rey 15 item test), language (phonological and semantic verbal fluency), executive function (Raven's matrices, Wisconsin card sorting test, frontal assessment battery), visuospatial skills (Rey-Osterrieth complex figure) (Carlesimo et al., 1996; Bianchi and Dai Prà, 2008). Raw scores were corrected for age and education and then transformed into equivalent scores, ranging from zero (pathological) to four (normal). For each cognitive domain, a weighted score was obtained averaging the value of the equivalent scores of all tests belonging to that specific cognitive domain (van Dijk et al., 2008). Clinical classification of AD or VD was performed according to the abovementioned criteria and was further refined by excluding patients with mixed dementia according to the Hachinski scale (HS) with pathology-validated cutoffs (Hachinski et al., 1975; Moroney et al., 1997): pure VD (HS  $\geq$  7), pure AD (HS  $\leq$  4), and MXD (HS 5 and 6). Vascular alterations were semi-quantitatively rated on radiological bases by evaluating white matter (WM) leukoaraiosis using the Fazekas scale (Fazekas et al., 1987).

## MRI Acquisitions

All subjects underwent MRI examination using a 3T Siemens Skyra scanner (Siemens, Erlangen, Germany) with a 32-channel head coil. The MRI acquisition protocol included (1) rs-fMRI: T2\*-weighted gradient echo echo-planar (GRE-EPI) sequence (TR/TE = 3,010/20 ms; voxel size = 2.5 mm isotropic, FOV = 224 mm, 60 slices, 120 volumes) and (2) DTI: twice refocused spin echo echo-planar (SE-EPI) sequence (TR/TE = 10,000/97 ms, 70 slices with no gap, acquisition matrix = 120 × 120, voxel size = 2 mm isotropic, 64 diffusion-weighted directions, b = 1,200 s/mm<sup>2</sup>, 10 volumes with no diffusion weighting). A high-resolution 3-D sagittal T1-weighted (3-D T1) scan (MPRAGE sequence: TR/TE/TI = 2,300/2.95/900 ms, flip angle 9°, 256 slices, voxel size = 1 × 1 × 1.2 mm<sup>3</sup>, FOV = 270 mm) was also acquired for anatomical reference.

## Image Processing and Data Analysis

Image analysis was carried out using the FSL tools (FMRIB Software Library, version 5.0.9, <http://www.fmrib.ox.ac.uk/fsl/>) and Matlab (v. R2018b, The Mathworks, Inc., Natick, MA).

## DTI Analysis

### Data Preprocessing

First, for each subject, the 10 volumes acquired with no diffusion weighting ( $b_0 = 0$  s/mm<sup>2</sup>) were averaged to obtain a mean  $b_0$  volume. DTI volumes were then corrected for motion and eddy current distortions using the eddy tool, which aligns the diffusion-weighted volumes to the mean  $b_0$  image. A binary brain mask was obtained from the mean  $b_0$  volume using the brain extraction tool (BET). DTIFIT was used to generate individual fractional anisotropy (FA) and mean diffusivity (MD) maps. For each subject, the 3-D T1 images were first segmented with the FAST algorithm of FSL to produce the WM and GM maps (as well as whole brain by their addition). The FA map was then aligned to the respective 3-D T1 volume using a full-affine registration with a windowed-sinc interpolation (using a Hanning window of size 7 × 7 × 7) using the FMRIB's linear image registration tool (FLIRT). The mean FA value of the brain was obtained by overlapping the brain mask with the aligned FA map.

### Tract-Based Spatial Statistics

Tract-based spatial statistics (TBSS) was performed on DTI images to investigate the voxel-wise distribution of FA and MD differences among groups. This analysis was carried out using the TBSS tool as implemented in FSL and following the pipeline reported in Smith et al. (2006). The FA maps of all the subjects were aligned to the best target and then to a common space (MNI152 space) by non-linear registration and averaged to obtain a mean FA skeleton. Finally, each subject's aligned FA data were projected onto the mean FA skeleton. MD maps were also projected into the mean FA skeleton using the same projector vectors that were obtained in the FA maps alignment. A GLM was applied to assess differences in FA and MD between patients and HC. The results of this study have been fully reported in Palesi et al. (2018).

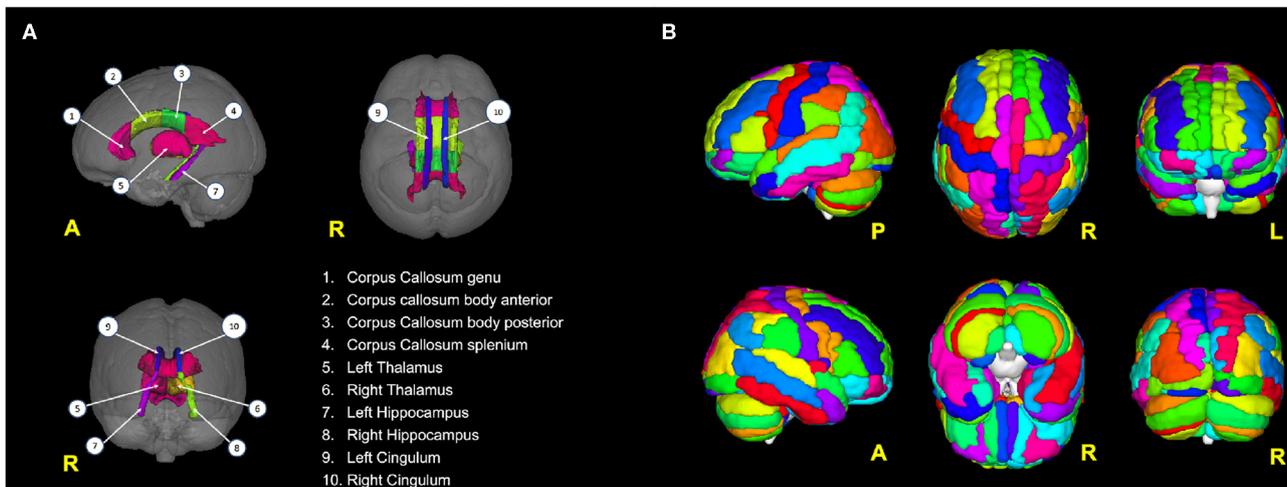
### Features Extraction From DTI

The areas that resulted in being particularly relevant from the TBSS analysis were saved as regions of interest (ROIs, see **Figure 1A** and **Table 1**). For each ROI, we then extracted mean FA and MD values. These extracted values were used as DTI-derived features for the ML approach of this study.

## rs-fMRI Analysis

### Data Preprocessing

Individual subject's preprocessing consisted of motion correction, brain extraction, spatial smoothing using a Gaussian kernel of full-width-at-half-maximum (FWHM) of 5 mm, and high-pass temporal filtering equivalent to 150 s (0.007 Hz). For each subject, rs-fMRI volumes were then registered to the corresponding structural 3-D T1w scan using the FMRIB's linear image registration tool (FLIRT) and subsequently to standard space (MNI152) using FMRIB's non-linear image registration tool (FNIRT) with default options. Moreover, to reduce the nuisance effects of non-neuronal BOLD fluctuations, white matter (WM) and cerebrospinal fluid (CSF) signals were regressed out from rs-fMRI data.



**FIGURE 1 | (A)** Brain areas from which DTI features (FA, MD) have been extracted; **(B)** the 116 areas from the AAL atlas used to parcellate the brain and then to calculate the rs-fMRI-derived graph theory (GT) metrics.

**TABLE 1 |** Overview of the imaging features used to form the three data sets: DTI, fMRI GT, and DT + fMRI GT data sets.

Dataset	ROI	Imaging features	N features
DTI	Corpus callosum (genu, body anterior, body posterior, splenium) thalamus (left, right) hippocampus (left, right) cingulum (left, right)	FA, MD	20
fMRI GT	116 AAL areas	Nodal degree (DEG) participation coefficient (PC) betweenness centrality (BC) clustering coefficient (CC) normalized local efficiency ( $E_{loc_{norm}}$ ) global efficiency (Eglob) graph average CC (Cm)	698
DTI + fMRI GT	Corpus callosum (genu, body anterior, body posterior, splenium) thalamus (left, right) hippocampus (left, right) cingulum (left, right) 116 AAL areas	FA, MD nodal degree (DEG) participation coefficient (PC) betweenness centrality (BC) clustering coefficient (CC) normalized local efficiency ( $E_{loc_{norm}}$ ) global efficiency (Eglob) graph average CC (Cm)	718

### Brain Network Computation

For each subject, preprocessed rs-fMRI images were parcellated using the automated anatomical labeling (AAL) atlas into 116 distinct areas (Tzourio-Mazoyer et al., 2002) that defined the *nodes* of the brain network (Figure 1B; see also Table S1). For each AAL area, the mean rs-fMRI signal was calculated by averaging the time series of all the voxels within the AAL region. The *edges* of the brain network were defined

as the functional connectivity of all pairs of 116 AAL areas using Pearson's correlation coefficient. This operation generated for each subject a weighted undirected functional connectivity matrix, which corresponded to a dense network. Each subject-specific connectivity matrix was then thresholded by preserving a proportion  $P$  ( $0 < P < 1$ ) of the strong weights, which corresponds to the number of the retained strong weights divided by the total number of weights. All diagonal weights (self-connections) were set to zero.  $P$  was set at 12% to obtain the optimal matrix sparsity as suggested by Rubinov and Sporns (2010).

### Graph Metrics Computation

Each graph (i.e., thresholded correlation matrix) was then treated in Matlab with the Brain Connectivity Toolbox (BCT) to compute different metrics from graph theory (GT) to investigate the properties of the 116 AAL brain regions. Specifically, three measures of functional segregation: the clustering coefficient (CC), the graph average CC (Cm), and the normalized local efficiency ( $E_{loc_{norm}}$ ) were calculated to characterize the ability of the brain for specialized processing to occur within a densely interconnected group of regions. A functional integration measure, global efficiency (Eglob), was used to assess the ability of the brain to rapidly combine specialized information from distributed regions. Moreover, three local nodal measures were calculated: nodal degree (DEG), participation coefficient (PC), and betweenness centrality (BC) (Rubinov and Sporns, 2010).

### Data Sets for Machine Learning

All MRI features were obtained from the baseline visit of the patients. For each subject, the measures extracted from DTI (i.e., mean FA and MD values from 10 selected brain areas) and rs-fMRI (graph metrics from 116 AAL areas) were combined in a vector of parameters (i.e., a record). The records of all subjects

were then collected to create the following data sets (see also **Table 1**):

1. DTI data set (unimodal data set): This contained only the DTI-derived metrics (20 features per subject). Therefore, it was considered a unimodal data set.
2. GT data set (unimodal data set): This contains the 698 graph theoretical metrics per subject, derived from rs-fMRI images. This was also considered a unimodal data set.
3. DTI + GT data set (multimodal data set): This was obtained unifying (1) and (2) into a single data set, which resulted in 718 features per subject.

Each data set was screened to remove outliers by deleting those records that contained more than 30% of features laying three standard deviations (*SD*) away from the sample's mean. The threshold of 3 *SD* was chosen as a good value to remove the spurious feature values, which definitively fell outside the 99.7% of their distributions, from the database (Rousseeuw and Hubert, 2011). Different thresholding percentages on the subjects' features were tested, varying between 10 and 80%, and the 30% threshold resulted in the best compromise between data cleaning and data preservation. Moreover, for each data set, data that survived the outlier-removal procedure were standardized (i.e., *z*-score normalized) prior to beginning the following steps of the analysis. In order to check whether multimodal features were superior to unimodal ones in separating the two patients' groups, each ML algorithm was run separately on three distinct data sets: DTI data set (unimodal), GT fMRI data set (unimodal), and DTI + fMRI GT data set (multimodal).

## Feature Selection

Not all imaging metrics are useful for classification because of their intrinsic redundancy. Feature selection is, therefore, important for extracting the most informative features for the specific task. Moreover, a high-dimensional data set may lead to over-fitting issues. For all these reasons, in this paper, the ReliefF feature selection algorithm was applied to each data set (see section Datasets for Machine Learning) prior to running any ML algorithm (Kononenko et al., 1997). This procedure produced a ranking of features according to their relevance in determining the class value of the data set records. Indeed, this step allowed us to select only the most informative features (and to discard the irrelevant ones) in order to improve the ML algorithms' performances. After the feature-selection stage, the ML algorithms identified as promising for the task of the study were used to construct the classification models. More details about how ReliefF was applied in our analyses have been fully reported in section Cross Validated Accuracy.

## Machine Learning Analysis

Three supervised ML approaches were considered for binary classification: artificial neural network (ANN), support vector machine (SVM), and adaptive neuro-fuzzy inference system (ANFIS). All the methods were implemented in Matlab as part of a dedicated image-based tool. The work was organized into two main steps as follows.

The selected ML algorithms (i.e., classifiers) were first run separately to

- *AIM 1* (model construction and validation): For each ML algorithm, a tuning of the relevant parameters was performed in order to identify the optimal setting to maximize the algorithm classification performance. We then proceeded by training the model and testing it independently using a balanced cross-validation approach (see section Cross Validated Accuracy). For each algorithm, for each data set (as in section Datasets for Machine Learning), we identified the best feature set to discriminate AD from VD as the one associated with the best classification performance.

Finally, we selected only the ML algorithm with the best classification performance in the binary task and we used it to

- *AIM 2* (prediction): We predicted the prevalent underlying disease in the MXD subjects, using as input the discriminant feature pattern previously identified by the selected algorithm during the training step.

## Artificial Neural Network (ANN)

ANN are a family of learning methods inspired by biological neural networks (Haykin, 1998). In this study, two different ANN models have been implemented: multilayer perceptron (MLP) (Rumelhart et al., 1986; Rumelhart and McClelland, 1987; Geva and Sitte, 1992) and radial basis function network (RBFN) (Acosta, 1995; Bishop, 2006).

*Multilayer perceptron (MLP)*: The MLP implementation for this study was performed in Matlab and was composed of three layers: an input layer with  $n$  nodes corresponding to the  $n$  input features from the calculated data set, a hidden layer with eight nodes and a one-node output layer. The output node resulted in either zero or one, respectively, for the AD or VD class. A sigmoidal activation function (*tansig*) was used to transform data between the input and the hidden layer as well as between the hidden layer and the output layer. A Bayesian regularization back-propagation approach was used to train the MLP network (Bishop, 2006).

*Radial Basis Function network (RBFN)*: This is a variant of the three-layer feed-forward neural network, which uses radial basis (Gaussian) functions as its activation functions (Bishop, 2006). In this study, the RBFN was implemented in Matlab using the *newrbe* function with the spread constant for the radial basis layer set equal to 0.1.

## Support Vector Machine (SVM)

SVM uses training data to find the maximal margin hyperplane that best divides data belonging to different groups or classes (Cortes and Vapnik, 1995). The separating hyperplane is selected to have the largest distance from the nearest training data points of any class. In the case of non-linearly separable data, a non-linear kernel function is used to project them into a higher dimensional space where they can be linearly separated. For the present study, two SVM architectures with a different non-linear kernel function were used: SVM with a radial basis function (RBF) kernel (SVM<sub>RBF</sub>) and SVM with an MLP sigmoid-like

kernel (SVM<sub>MPL</sub>). For each SVM architecture, an iterative grid search was performed in order to find the optimal combination of  $C$ ,  $\sigma$  (for SVM<sub>RBF</sub>),  $\alpha$ , and  $c$  (for SVM<sub>MPL</sub>) to obtain the best SVM performance with our data.

### Adaptive Neuro-Fuzzy Inference System (ANFIS)

ANFIS is a class of ANN that represents a trade-off between ANN and fuzzy logic systems, offering smoothness due to the fuzzy control interpolation and adaptability due to the ANN back-propagation (Zadeh, 1978). ANFIS incorporates both ANN and fuzzy logic principles and converges the benefits from both the methods in a single implementation. For this work, the ANFIS algorithm has been used as implemented in the Fuzzy Logic toolbox in Matlab using a Sugeno-type fuzzy inference system (FIS) and Gaussian functions as membership functions to specify the fuzzy sets. A hybrid learning algorithm, obtained by combining the least-squares and back-propagation gradient descent methods, was used to model the training data. For the purposes of the study, ANFIS was run using 100 epochs.

The ML models implemented for the present study can be considered as variants of the analog deep learning algorithms, which differ from the ML ones for their ability to learn features automatically at multiple levels, therefore allowing the system to learn complex functions mapping the input to the output directly from data (Goodfellow et al., 2016).

### Cross-Validated Accuracy

To improve the classification performance robustness of each ML method, for AIM1, we adopted a balanced Monte Carlo 10-fold cross-validation (CV) approach using 100 bootstraps. Indeed, at each iteration, the CV algorithm divided the original input data into 10 parts with the two classes (AD and VD) equally represented. This operation resulted in the creation of 100 new different CV data sets. Specifically, for each CV data set, for each ML algorithm, nine parts (i.e., 9-folds) were used as training subset, and the remaining one part (i.e., 1-fold) was used as testing subset. Moreover, the ReliefF feature-selection algorithm described above (see section Feature Selection) was here applied to each generated CV data set. For each classifier, we considered the best classification performance, obtained over the 100 bootstraps, and its related model and associated selected features as final results. This approach allowed reduction variance of the data, therefore reducing the chance of over-fitting or bias errors.

### Performance Comparison

In this study, an AD subject effectively classified as AD was considered a true positive (TP), and a VD subject effectively classified as VD was counted as a true negative (TN). An AD subject classified as VD was considered a false negative (FN), and a VD subject classified as AD was counted as a false positive (FP). Given that, for each experiment, the classification performance of the constructed model, varying the algorithm parameters as well as the number of input features presented to it, was assessed by calculating the *classification accuracy* ( $ACC = (TP + TN)/(TP + TN + FP + FN)$ ), which denotes the probability of a correct classification; *sensitivity* ( $SEN = TP/(TP + FN)$ ), which scores

the ability of the model to detect a subject with a specific disease in a population with more than one disease; *specificity* ( $SPE = TN/(TN + FP)$ ), which scores the ability of the model to correctly rule out the disease in a disease-free population; *precision* ( $PREC = TP/(TP + FP)$ ), which defines the proportion of positive predictions; and *negative predictive value* ( $NPV = TN/(TN + FN)$ ), which denotes the proportion of negative predictions (Bishop, 2006). For SVMs, the mean ratio of support vectors (SVr), calculated as the number of support vectors divided by the number of training subjects, was also reported as a measure of complexity degree of the models. The receiver operating characteristic (ROC) curve was calculated for each implemented model, and its area under the curve (AUC) was used to compare the different classifiers' performance (Hanley and McNeil, 1982). For each performance score, the 95% confidence intervals (95% IC) were computed using the Wilson score interval with the continuity correction approach (Newcombe, 1998).

### Prediction on MXD

The ML model that showed the highest classification performance in AIM1 was then considered to fulfill the prediction purposes of AIM2. To achieve this task, we considered the AD-VD discriminant feature pattern that resulted from the best performing classifier in AIM1. We then used this feature pattern as input for the selected ML algorithm to predict the prevalent underlying disease (i.e., AD or VD) in the MXD subjects. Finally, for each MXD subject, the predicted class from the ML algorithm on baseline MRI data was compared with the patient's 3-year follow-up clinical evaluation in order to assess the reliability of the ML predictions as well as the potential of ML to notify earlier (than clinical evidence) about the typology of the patient's dementia.

### Non-imaging Statistics

Statistical analyses were carried out using SPSS (version 21.0; SPSS, Chicago, IL, USA). Demographic, behavioral, and radiological continuous data were first tested for normality using the Shapiro-Wilk test, and differences between groups were assessed with different tests depending on the typology of the variables (binary, normally or non-normally distributed). Chi-square tests were performed to compare frequency distributions of gender in the three groups. One-way analysis of variance (ANOVA) with Bonferroni correction was used to assess whether age was statistically different between groups (AD, VD, and MXD). A non-parametric Kruskal-Wallis test was applied to test differences between the groups in education level, clinical indices (HS, Fazekas and AWMRC, see section Clinical and Neuropsychological Assessment for details), and neuropsychological scores (attention, memory, language, executive and visuospatial cognitive domains). A non-parametric Mann-Whitney  $U$ -test was performed to test differences between paired groups in HS and Fazekas. A further Mann-Whitney  $U$ -test was applied to test differences in the features that the most performing ML algorithm identified as relevant to separate AD and VD.

**TABLE 2** | Demographics of AD, VD, and MXD groups.

	AD	VD	MXD	<i>p</i>
N	33	27	15	
Sex (M/F)	15/18	3/24 <sup>†</sup>	7/8 <sup>†</sup>	<b>&lt;0.05*</b>
Age	72.88 ± 7.31	76.67 ± 7.77	76.33 ± 6.78	0.123
Education (years)	6.47 ± 3.26	5.52 ± 2.13	5.08 ± 1.65	0.363
MMSE	16.10 ± 6.32	17.89 ± 4.15	18.59 ± 4.17	0.245
Memory	0.65 ± 0.74	0.73 ± 0.61	0.64 ± 0.70	0.832
Attention	0.91 ± 0.90	0.71 ± 0.68	0.48 ± 0.53	0.185
Language	1.06 ± 1.26	1.02 ± 1.06	1.08 ± 1.48	0.986
Executive function	0.60 ± 0.91	0.45 ± 0.76	0.38 ± 0.87	0.665
Visuospatial skills	0.66 ± 1.41	0.52 ± 1.19	0.30 ± 1.25	0.682
Hachinski score (HS)	2.97 ± 0.84 <sup>†</sup>	8.27 ± 1.58 <sup>†</sup>	5.80 ± 2.11 <sup>†</sup>	<b>&lt;0.001*</b>
Fazekas score	2.65 ± 1.27 <sup>†</sup>	4.63 ± 1.52 <sup>†</sup>	4.21 ± 1.82 <sup>†</sup>	<b>&lt;0.001*</b>

Demographic and clinical scores for Alzheimer's disease (AD), vascular dementia (VaD), and mixed VD-AD dementia (MXD) groups. Values are expressed as mean ± SD. MMSE, Mini Mental State Examination. *p*-values show statistically significant differences between AD, VD and MXD groups. \**p* < 0.05 between AD and VD. <sup>†</sup>*p* < 0.05 between MXD and AD or VD or between AD and VD. Statistically significant *p*-values have been highlighted in bold.

**TABLE 3** | Classification performance scores obtained using data, respectively, from the DTI, GT fMRI, and DTI + GT data sets.

	ACC (%)	SENS (%)	SPEC (%)	PREC (%)	NPV (%)
<b>DTI DATASET</b>					
SVM <sub>RBF</sub>	79.75 (66,89)	68.00 (54,79)	91.50 (80,97)	88.89 (76,95)	88.89 (77,95)
SVM <sub>MLP</sub>	73.00 (59,84)	63.00 (49,75)	83.00 (70,92)	78.75 (65,88)	79.00 (65,88)
MLP	75.00 (61,85)	74.00 (60,85)	76.00 (62,86)	75.51 (62,86)	74.50 (60,85)
RBFN	60.25 (46,73)	65.00 (51,77)	55.50 (42,69)	59.36 (45,72)	38.34 (25,52)
ANFIS	<b>83.50 (71,92)</b>	<b>81.00 (68,90)</b>	<b>86.00 (74,94)</b>	<b>85.26 (73,93)</b>	<b>79.90 (66,89)</b>
<b>GT fMRI DATASET</b>					
SVM <sub>RBF</sub>	81.00 (68-90)	93.50 (83,99)	68.50 (55,80)	74.80 (61,85)	74.80 (61,85)
SVM <sub>MLP</sub>	78.25 (64-88)	81.00 (68,90)	75.50 (62,86)	76.78 (63,86)	76.78 (63,87)
MLP	58.25 (44,71)	55.50 (42,69)	61.00 (47,74)	58.73 (45,72)	57.81 (43,70)
RBFN	55.75 (42,69)	55.50 (42,69)	56.00 (42,69)	55.78 (42,69)	17.06 (8,29)
ANFIS	<b>82.75 (69,91)</b>	<b>73.50 (64,88)</b>	<b>92.00 (81,97)</b>	<b>90.18 (78,96)</b>	<b>73.60 (59,84)</b>
<b>DTI+GT DATASET</b>					
SVM <sub>RBF</sub>	84.75 (72,93)	84.00 (71,92)	85.50 (73,93)	85.28 (73,93)	85.28 (72,93)
SVM <sub>MLP</sub>	74.75 (61,85)	65.00 (51,77)	84.50 (72,93)	80.75 (68,90)	80.75 (67,89)
MLP	76.75 (63,87)	74.00 (60,85)	79.50 (66,89)	78.31 (65,88)	75.35 (61,85)
RBFN	62.75 (49,75)	86.50 (74,94)	39.00 (27,53)	58.64 (45,71)	74.28 (60,84)
ANFIS	<b>85.25 (73,93)</b>	<b>82.00 (69,91)</b>	<b>88.50 (77,96)</b>	<b>87.70 (76,95)</b>	<b>81.69 (68,90)</b>

Each performance score is expressed as a percentage (%). For each score, the 95% IC is also reported in brackets. Classification performance scores (accuracy, sensitivity, specificity and AUC) for pairwise classifiers are expressed as percentage (%) with 95% confidence intervals in brackets. AUC, Area under receiver operating characteristic curve. ANFIS emerged as the most performant method in discriminating at baseline AD from VD independently of the data set (scores highlighted in bold).

## RESULTS

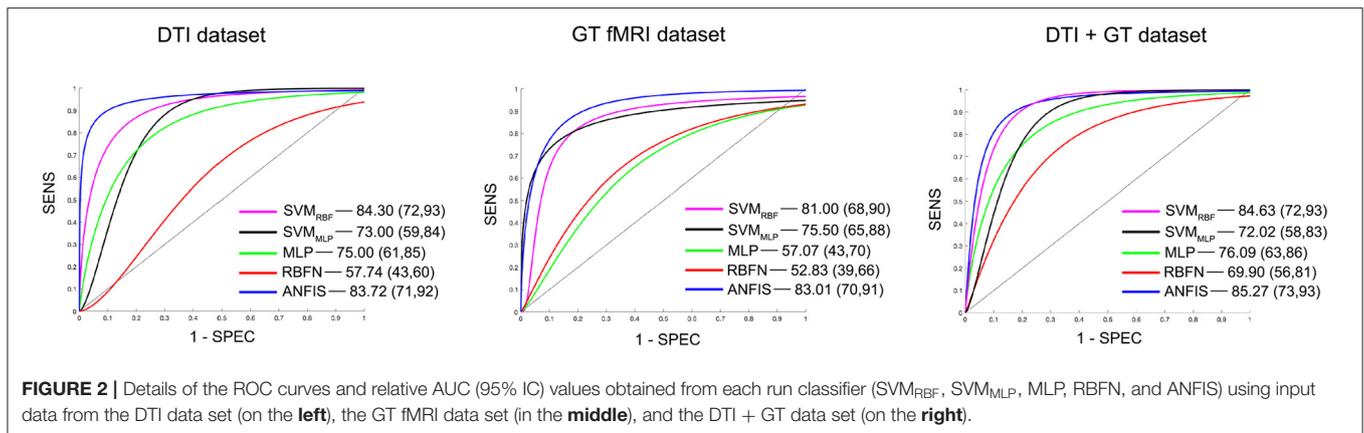
### Clinical Findings

The demographic and clinical characteristics of patients are summarized in **Table 2**. Significant differences were found in gender between AD and VD as well as between VD and MXD groups. Fazekas and HS scores showed significant differences between AD and VD. Significant differences were also found in HS when comparing MXD vs. AD and VD groups. Fazekas scores were also significantly higher in MXD compared to AD.

### Classification Results

Three different kinds of ML algorithms (SVM, ANN, and ANFIS) were used to identify the best feature pattern to classify AD from VD. The analyses yielded the following results (details of the classification performances of each classifier on the three data sets have been fully reported in **Table 3**, **Figure 2** and **Table S2**).

*Classification using the DTI data set:* ANFIS showed the best classification performance in dividing AD from VD subjects, reaching a classification accuracy (ACC) of 83.50% and area under the ROC curve (AUC) equal to 83.72% (see **Table 3**



**TABLE 4 |** Details (number of features, brain area, and MRI metric) of the discriminant feature pattern identified by ANFIS, which resulted the best classifier in separating AD and VD using data from the multimodal data set (i.e., DTI + GT data set).

	ACC (%)	N features	Area	MRI metric	Relieff ranking score
ANFIS	85.25	10	L Thalamus	FA	0.384
			Corpus callosum body anterior	FA	0.312
			R Anterior Cingulum	DEG	0.299
			Corpus callosum genu	FA	0.284
			L Precuneus	DEG	0.268
			L Hippocampus	FA	0.255
			R Superior Parietal gyrus	DEG	0.206
			L Fusiform gyrus	DEG	0.202
			Corpus callosum body posterior	FA	0.172
			R Fusiform gyrus	DEG	0.149

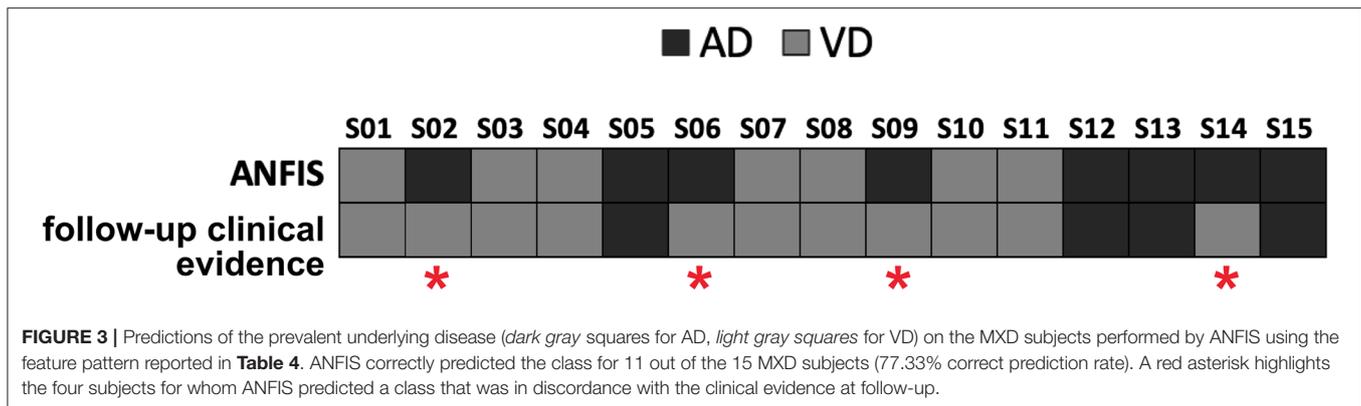
The listed features are reported reflecting the ranking order given by ReliefF (i.e., top feature corresponds to the most informative one). For each feature, the ReliefF ranking score has also been added. L, left part; R, right part 2.

and **Figure 2**), using a feature pattern including only four FA features from four brain regions, including the left hippocampus and three areas of the corpus callosum (body anterior, genu, and splenium, see **Table S2**). SVM<sub>RBF</sub> ( $C = 1$ ,  $\sigma = 1.87$ , SVr = 78.57%) also discriminated the two patient groups with a relatively high ACC = 79.75% and AUC = 84.30%, using a feature pattern of seven FA features from seven distinct brain regions including both hippocampi, left cingulum, left thalamus, and the corpus callosum (body anterior/posterior, genu, see **Table S2**). The remaining ML algorithms (SVM<sub>MLP</sub>, MLP, and RBFN) resulted instead in lower classification performance (ACC  $\leq$  75%) as reported in **Table 3**.

**Classification using the GT fMRI data set:** ANFIS reached the highest performance score with ACC = 82.75%, AUC = 83.01% using 10 features (mainly DEG and Eloc<sub>norm</sub>) from nine distinct brain areas, including bilateral superior parietal gyrus, right anterior cingulum, left precuneus and cuneus, left superior frontal gyrus, right postcentral gyrus, and bilateral fusiform gyri (**Table 3**, **Figure 2** and **Table S2**). SVM<sub>RBF</sub> ( $C = 0.14$ ,  $\sigma = 1.73$ , SVr = 98.21%) resulted in a relatively lower (than ANFIS) performance, scoring ACC = 81% and AUC = 81% using six features (mainly DEG and Eloc<sub>norm</sub>) from six distinct brain regions, including left precuneus and cuneus, right middle

and left superior frontal gyri, right postcentral gyrus, and right fusiform gyrus (**Table S2**). Moreover, SVM<sub>MLP</sub> ( $C = 0.37$ ,  $\alpha = 1$ ,  $c = -1$ , SVr = 80.36%) showed a classification performance closer (even lower) to those of ANFIS and SVM<sub>RBF</sub>, reaching ACC = 78.25% and AUC = 75.50% (**Table 3** and **Figure 2**) using only three features (DEG and Eglob) from two distinct brain areas: right fusiform and superior parietal gyri (**Table S2**). The remaining ANN algorithms (MLP and RBFN) resulted, instead, in poorer classification performance (ACC < 60%) as reported in **Table 3**.

**Classification using the DTI + GT data set:** Even when using the multimodal data set (DTI + GT), ANFIS showed the best performance (**Table 3** and **Figure 2**), scoring ACC = 85.25% and AUC = 85.27% using a total of 10 features, including five FA features from DTI involving left hippocampus, left thalamus, and corpus callosum (genu, body anterior/posterior) and five features from GT fMRI, involving right anterior cingulum, right superior parietal gyrus, left precuneus, and bilateral fusiform gyri (see **Table 4** and **Table S2**). After ANFIS, SVM<sub>RBF</sub> ( $C = 2.72$ ,  $\sigma = 2.12$ , SVr = 91.07%) resulted in the second most performant classifier, reaching ACC = 84.75% and AUC = 84.63%, using a total of nine features, including four FA features from DTI involving left hippocampus, left thalamus, and corpus callosum (genu,



body anterior); and five features from GT fMRI, involving right anterior cingulum, right superior parietal gyrus, left precuneus, left cuneus, and right precentral gyrus (Table S2). The remaining algorithms (SVM<sub>MLP</sub>, MLP, and RBFN) showed better results when compared to the scores obtained using the unimodal data sets but lower (ACC < 80%) than ANFIS performance overall (see Table 3).

## Prediction Results

ANFIS resulted in the most performant classifier in separating AD from VD subjects (see Table 3), and its performance was maximized when using feature patterns that included multimodal features (i.e., combined DTI and GT features; see Table 4). Therefore, ANFIS was chosen to perform the subsequent analysis (i.e., AIM2 analysis). To achieve our AIM2 goal (prediction), the discriminant 10-feature pattern identified at AIM1 (see Table 4) was used as input to ANFIS to make predictions on the prevalent underlying disease (i.e., AD or VD) in MXD subjects. According to the evidence of patients' diagnosis from clinical evaluation at 3-year follow-up, ANFIS correctly predicted the prevalent underlying disease for 11 of the 15 MXD subjects, which means a 77.33% correct prediction rate (see Figure 3).

## DISCUSSION

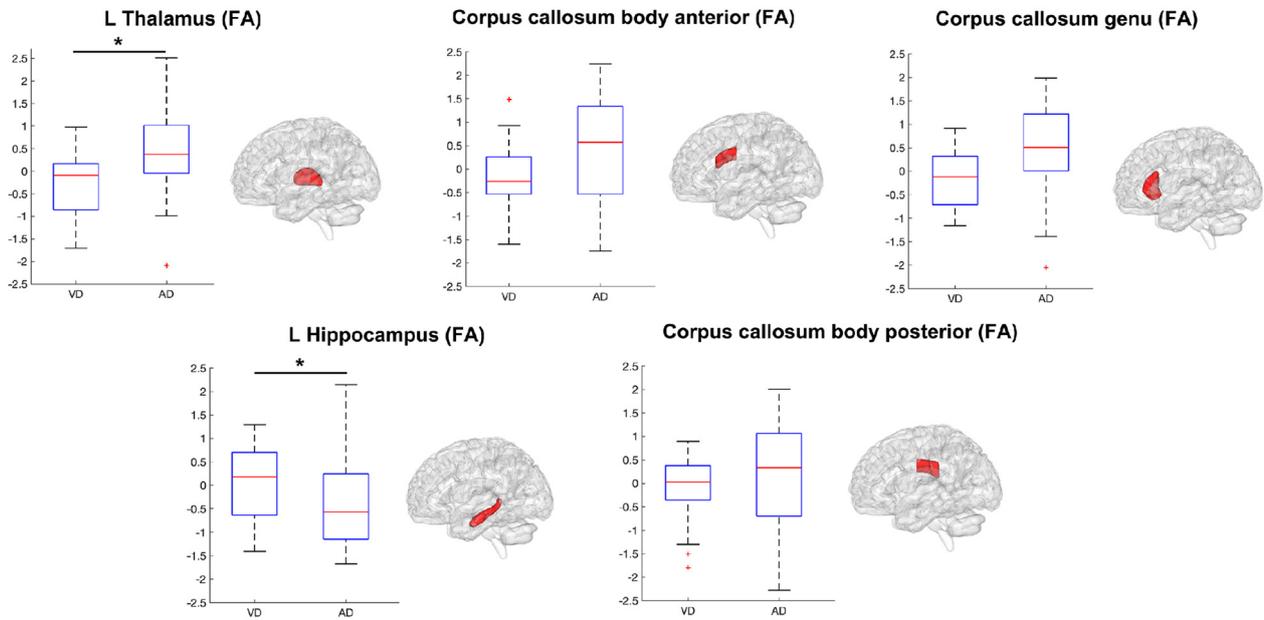
In this paper, ML algorithms were combined with advanced qMRI metrics to assess their potential to automatically discriminate AD from VD. First, quantitative metrics from DTI (FA and MD) and rs-fMRI (GT metrics) were extracted and used to build two unimodal data sets (DTI data set, GT fMRI data set). Then, the two data sets were unified in order to obtain a multimodal structural-functional data set (DTI + GT data set). Multiple supervised ML algorithms were applied on each data set, and classification results were obtained. Various ML methods have been used in literature to identify dementia-like diseases, including SVM with different kernels and ANNs, such as RBFN, MLP, and ANFIS. We examined these classifiers to find the most performant for our study. Finally, the best discriminant feature pool was identified and used as input for the most performant classifier to predict the prevalent underlying disease (i.e., AD rather than VD) on a group of subjects whose

diagnosis, based on clinical evaluation, was unclear and defined as mixed between VD and AD (MXD group). The algorithm's accuracy in predicting the MXD prevalent underlying disease was compared with the diagnosis evidence obtained from 3-year follow-up clinical screening.

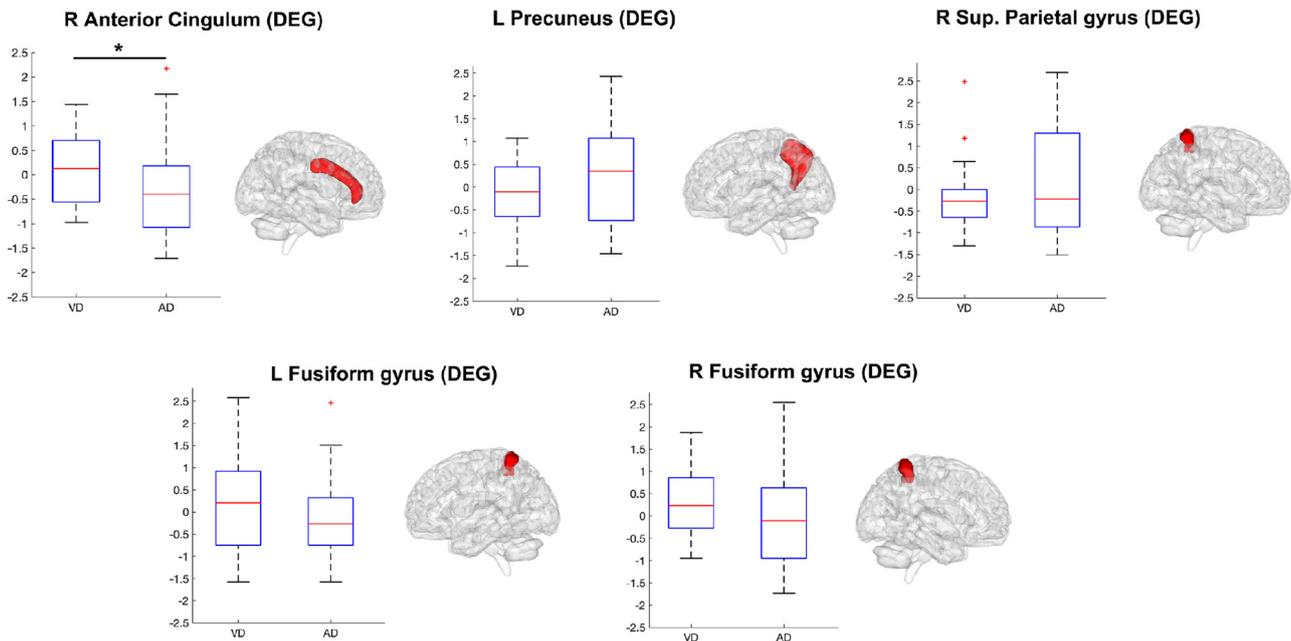
Among the evaluated algorithms, ANFIS emerged as the most performant method in discriminating at baseline AD from VD independently of the data set (unimodal or multimodal) used as input (see Table 3). Indeed, ANFIS achieved the highest classification accuracy (ACC > 84%) when using the multimodal data set as input, i.e., when providing both structural and functional information as input, simultaneously. These results are in line with the findings of a number of previous studies that investigated the use of multimodal data to automatically diagnose AD (and MCI) from healthy subjects (Zhang et al., 2011; Liu F. et al., 2014; Liu M. et al., 2014; Lei et al., 2016; Liu et al., 2018). All these studies concluded that unimodal data generally provides incomplete information to accurately diagnose dementia, and multimodal data tends to boost the classification accuracy due to the complementary information.

Table 4 lists the pattern of features that ANFIS identified as the most discriminant to separate AD from VD. Brain areas such as the thalamus, hippocampus, precuneus, and anterior cingulum were identified as relevant to discriminate AD and VD subjects. Because AD dementia involves a significant deficit in the anterograde episodic memory, areas such as the hippocampus and the thalamus were expected to come out as relevant for the classification problem (Jahn, 2013; Aggleton et al., 2016). The anterior cingulum and the precuneus are both relevant for episodic memory as well and are also core regions of the default mode network (DMN), which has been extensively studied because of its severe alterations in AD (Greicius et al., 2004; Castellazzi et al., 2014). According to the implemented feature selection algorithm (ReliefF), the FA of the left thalamic white matter resulted as the most informative feature to discriminate AD from VD (Table 4). The thalamus region is involved in neural networks that sustain complex cognitive and behavioral functions, and a link has been demonstrated between thalamic dysfunctions and episodic memory impairment in AD (Aggleton et al., 2016). Indeed, our results revealed that the left thalamic FA was significantly higher ( $p < 0.05$ ) in AD compared to VD (see Figure 4), therefore suggesting that, in AD, the thalamic

## White Matter features



## Grey Matter features



**FIGURE 4 |** Boxplots representing the summary of the 10 features (WM features on top, GM features on bottom) in AD and VD groups. The ensemble of these features (see also **Table 4**) formed the discriminant pattern that was also used to predict the prevalent underlying disease in MXD subjects. Each feature has been tested with the Mann–Whitney  $U$  test in order to assess significant differences between AD and VD values. An asterisk mark has been added on the top of the boxplot of the features with values significantly ( $p < 0.05$ ) different between the AD and VD populations.

tracts might be more coherent possibly due to a loss of crossing fibers (Teipel et al., 2014). Moreover, the anterior cingulum nodal degree (DEG), which reflects the number of functional connections between ACC and other parcellated brain areas, also resulted as an informative feature to discriminate AD from VD (Table 4). Indeed, the AD group showed significantly ( $p < 0.05$ ) lower DEG scores than VD (see Figure 4) in the anterior cingulum, therefore suggesting a more severe disconnection of this area in AD patients. In our study, the left hippocampal white matter FA, extracted by the TBSS analysis (Palesi et al., 2018), emerged as a relevant feature to separate AD from VD subjects (Table 4). Indeed, the FA in the left hippocampal area resulted significantly ( $p < 0.05$ ) reduced in AD compared to VD (see Figure 4). This result may be interpreted as the evidence of more severe tract disconnection affecting the WM portion of the left hippocampus in AD.

When using the AD–VD discriminant feature pattern (Table 4) as input to ANFIS to predict the prevalent underlying disease in MXD patients starting from their baseline data, the ML algorithm achieved a correct prediction rate of 77.33% (Figure 3). The accuracy of ANFIS predictions has been validated against the diagnosis obtained from MXD patients' clinical screening at 3-year follow-up. The high matching rate between ANFIS predictions and clinical follow-up suggests the potential of the ML approach combined with MRI-derived indexes to obtain an accurate detection of the prevalent disease in individual subjects with MXD 3 years prior to clinical evidence. Considering that novel therapeutic approaches to treat the AD condition, such as those using monoclonal antibodies, are more effective at early stages of the pathology (van Dyck, 2018), the ML-based prediction of AD in subjects clinically diagnosed with MXD could be crucial in order to promptly administrate these emerging treatments.

From the point of view of study design, the present work is a cross-sectional investigation, and although the ML algorithms responded with high performance in discriminating clinical AD vs. VD, their implication for MXD prognosis and their integration in patients' management will require appropriate longitudinal data. From a technical perspective, this study used ML to disentangle the feature pattern that better identifies the AD profile and separate it from the VD one, suggesting a possible solution to identify the most likely disease progression in subjects with MXD. Indeed, this will help to identify the most suitable and prompt therapy for each MXD subject.

A different approach for classification, which is becoming very popular, is to use deep learning algorithms, which differ from ML for their ability to learn features automatically at multiple levels, therefore allowing the system to learn complex functions mapping the input to the output directly from data (Goodfellow et al., 2016; Qureshi et al., 2019). It could be envisaged that ML and deep learning methods could lead beyond current clinical diagnosis by establishing, in an unsupervised fashion, groups of patients with similar MRI and clinical scores. This may lead beyond current clinical classification of dementia and require a major clinical effort to understand the biological mechanisms that differentiate potentially novel disease patterns, but this is beyond the scope of this study.

In this study, we used the AAL atlas to parcellate the brain before applying GT. The AAL atlas parcels the brain on the anatomical traits, which do not exactly match functional brain organization, and this may degrade performance metrics. Indeed, Shirer et al. (2012) showed that an atlas based on functional (rather than structural) ROIs provides better classification performances for the analysis of fMRI data. Nonetheless, there is still no consensus about the optimal strategy for brain parcellation (Arslan et al., 2018).

A final consideration is that the 85% accuracy of ANFIS was achieved solely based on qMRI features without including clinical tests. This means that objective qMRI features are able to perform the AD vs. VD classification alone, suggesting that this ML approach provides a substantial contribution for diagnosis. Future works should explore the pattern of features identified here together with clinical and neuropsychological variables and metrics from biological samples to improve the accuracy of the algorithm even further.

## CONCLUSIONS

This study, which combines local DTI metrics and GT measures from rs-fMRI data with ML, shows great potential for the automatic classification of AD and VD in patients with mixed clinical assessment. Indeed, multimodal features from MRI could be used to automatically separate AD from VD patients with high accuracy and balanced sensitivity and specificity. Among the pool of ML algorithms available to the user, ANFIS appeared to overcome others in classification performance. Results were consistent with reported literature in identifying specific brain regions such as the thalamus, hippocampus, and anterior cingulum with specific dementia types. Interestingly, our analytical method, by using baseline data, provided early prediction of disease type (AD or VD) in patients with clinical mixed dementia symptoms. Considering these encouraging results, we strongly believe that ML coupled to high-resolution MRI will provide a suitable approach to support clinicians in their clinical work, helping them to improve their diagnostic and prognostic accuracy as well as therapy and patient management.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Comitato Etico Pavia, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

GC, MGC, DM, GMa, and CG conceptualized the study. GC designed and performed the analyses with support AR and FP.

PV and NA acquired all MRI data. MCR, SB, FD, ES, AC, and GMi enrolled patients and acquired all the non-imaging data with the help of PV and NA. GMa, CG, DM, and MCR provided support and guidance with data interpretation with the clinical contribution of all physicians. GC, ED'A, GMa, and CG wrote the manuscript, with comments from all other authors.

## FUNDING

This work was supported by the Italian Ministry of Health (RC2016). CG received research funding from EPSRC, UK MS Society, Horizon2020, Biogen Idec, Novartis, Wings for Life. This study was funded by the UK MS Society (programme

grant number 984) and the ECTRIMS Postdoctoral Research Fellowship Programme to GC.

## ACKNOWLEDGMENTS

We thank Mr. Giancarlo Germani and Miss Arianna Faggioli for collaboration in data acquisition.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fninf.2020.00025/full#supplementary-material>

## REFERENCES

- Acosta, F. M. A. (1995). Radial basis function and related models: an overview. *Signal Process.* 45, 37–58 doi: 10.1016/0165-1684(95)00041-B
- Aggleton, J. P., Pralus, A., Nelson, A. J., and Hornberger, M. (2016). Thalamic pathology and memory loss in early Alzheimer's disease: moving the focus from the medial temporal lobe to Papez circuit. *Brain* 139(Pt 7), 1877–1890. doi: 10.1093/brain/aww083
- Agosta, F., Galantucci, S., and Filippi, M. (2017). Advanced magnetic resonance imaging of neurodegenerative diseases. *Neurol. Sci.* 38, 41–51. doi: 10.1007/s10072-016-2764-x
- Arslan, S., Ktena, S. I., Makropoulos, A., Robinson, E. C., Rueckert, D., and Parisot, S. (2018). Human brain mapping: a systematic comparison of parcellation methods for the human cerebral cortex. *Neuroimage* 170, 5–30. doi: 10.1016/j.neuroimage.2017.04.014
- Baskys, A., and Hou, A. C. (2007). Vascular dementia: pharmacological treatment approaches and perspectives. *Clin. Interv. Aging* 2, 327–335.
- Bianchi, A., and Dai Prà, M. (2008). Twenty years after Spinnler and Tognoni: new instruments in the Italian neuropsychologist's toolbox. *Neurol. Sci.* 29, 209–217. doi: 10.1007/s10072-008-0970-x
- Bishop, C. (2006). *Pattern Recognition and Machine Learning*. New York, NY: Springer Nature.
- Buckley, R. F., Schultz, A. P., Hedden, T., Papp, K. V., Hanseuw, B. J., Marshall, G., et al. (2017). Functional network integrity presages cognitive decline in preclinical Alzheimer disease. *Neurology* 89, 29–37. doi: 10.1212/WNL.0000000000004059
- Carlesimo, G. A., Caltagirone, C., and Gainotti, G. (1996). The Mental Deterioration Battery: normative data, diagnostic reliability and qualitative analyses of cognitive impairment. The Group for the Standardization of the Mental Deterioration Battery. *Eur. Neurol.* 36, 378–384. doi: 10.1159/000117297
- Castellazzi, G., Palesi, F., Casali, S., Vitali, P., Sinfioriani, E., Wheeler-Kingshott, C. A., et al. (2014). A comprehensive assessment of resting state networks: bidirectional modification of functional integrity in cerebro-cerebellar networks in dementia. *Front. Neurosci.* 8:223. doi: 10.3389/fnins.2014.00223
- Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297.
- Dallora, A. L., Eivazzadeh, S., Mendes, E., Berglund, J., and Anderberg, P. (2017). Machine learning and microsimulation techniques on the prognosis of dementia: a systematic literature review. *PLoS ONE* 12:e0179804. doi: 10.1371/journal.pone.0179804
- Dillen, K. N. H., Jacobs, H. I. L., Kukulja, J., Richter, N., von Reutern, B., Onur, Ö., et al. (2017). Functional disintegration of the default mode network in prodromal Alzheimer's disease. *J. Alzheimers Dis.* 59, 169–187. doi: 10.3233/JAD-161120
- Dyrba, M., Grothe, M., Kirste, T., and Teipel, S. J. (2015). Multimodal analysis of functional and structural disconnection in Alzheimer's disease using multiple kernel SVM. *Hum. Brain Mapp.* 36, 2118–2131. doi: 10.1002/hbm.22759
- Fazekas, F., Chawluk, J. B., Alavi, A., Hurtig, H. I., and Zimmerman, R. A. (1987). MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. *AJR Am. J. Roentgenol.* 149, 351–356. doi: 10.2214/ajr.149.2.351
- Filippi, M., Spinelli, E. G., Cividini, C., and Agosta, F. (2019). Resting state dynamic functional connectivity in neurodegenerative conditions: a review of magnetic resonance imaging findings. *Front. Neurosci.* 13:657. doi: 10.3389/fnins.2019.00657
- Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6
- Geva, S., and Sitte, J. (1992). A constructive method for multivariate function approximation by multilayer perceptrons. *IEEE Trans. Neural Netw.* 3, 621–624. doi: 10.1109/72.143376
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. Cambridge, MA: The MIT Press.
- Greicius, M. D., Srivastava, G., Reiss, A. L., and Menon, V. (2004). Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. *Proc. Natl. Acad. Sci. U.S.A.* 101, 4637–4642. doi: 10.1073/pnas.0308627101
- Groves, W. C., Brandt, J., Steinberg, M., Warren, A., Rosenblatt, A., Baker, A., et al. (2000). Vascular dementia and Alzheimer's disease: is there a difference? A comparison of symptoms by disease duration. *J. Neuropsychiatry Clin. Neurosci.* 12, 305–315. doi: 10.1176/jnp.12.3.305
- Hachinski, V. C., Iliff, L. D., Zilhka, E., Du Boulay, G. H., McAllister, V. L., Marshall, J., et al. (1975). Cerebral blood flow in dementia. *Arch. Neurol.* 32, 632–637
- Hanley, J. A., and McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143, 29–36. doi: 10.1148/radiology.143.1.7063747
- Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation*. Upper Saddle River, NJ: Prentice Hall PTR.
- Jahn, H. (2013). Memory loss in Alzheimer's disease. *Dialogues Clin. Neurosci.* 15, 445–454.
- Kononenko, I., Šimec, E., and Robnik-Šikonja, M. (1997). Overcoming the Myopia of Inductive Learning Algorithms with RELIEFF. *Appl. Intell.* 7, 39–55. doi: 10.1023/a:1008280620621
- Lei, B., Chen, S., Ni, D., and Wang, T. (2016). Discriminative learning for Alzheimer's disease diagnosis via canonical correlation analysis and multimodal fusion. *Front. Aging Neurosci.* 8:77. doi: 10.3389/fnagi.2016.00077
- Liu, F., Wee, C. Y., Chen, H., and Shen, D. (2014). Inter-modality relationship constrained multi-modality multi-task feature selection for Alzheimer's Disease and mild cognitive impairment identification. *Neuroimage* 84, 466–475. doi: 10.1016/j.neuroimage.2013.09.015
- Liu, M., Zhang, D., Shen, D., and Alzheimer's Disease Neuroimaging Initiative (2014). Hierarchical fusion of features and classifier decisions for Alzheimer's disease diagnosis. *Hum. Brain Mapp.* 35, 1305–1319. doi: 10.1002/hbm.22254

- Liu, X., Chen, K., Wu, T., Weidman, D., Lure, F., and Li, J. (2018). Use of multimodality imaging and artificial intelligence for diagnosis and prognosis of early stages of Alzheimer's disease. *Transl. Res.* 194, 56–67. doi: 10.1016/j.trsl.2018.01.001
- Long, X., Chen, L., Jiang, C., Zhang, L., and Alzheimer's Disease Neuroimaging Initiative (2017). Prediction and classification of Alzheimer disease based on quantification of MRI deformation. *PLoS ONE* 12:e0173372. doi: 10.1371/journal.pone.0173372
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., 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. *Alzheimers Dement.* 7, 263–269. doi: 10.1016/j.jalz.2011.03.005
- Miceli, G. (2006). Vascular dementia. *Neurol. Sci.* 27(Suppl. 1), S37–S39. doi: 10.1007/s10072-006-0545-7
- Moroney, J. T., Bagiella, E., Desmond, D. W., Hachinski, V. C., and Mölsä, P. K., Gustafson, L., et al. (1997). Meta-analysis of the Hachinski Ischemic Score in pathologically verified dementias. *Neurology* 49, 1096–1105.
- Newcombe, R. G. (1998). Interval estimation for the difference between independent proportions: comparison of eleven methods. *Stat. Med.* 17, 873–890.
- Palesi, F., De Rinaldis, A., Vitali, P., Castellazzi, G., Casiraghi, L., Germani, G., et al. (2018). Specific patterns of white matter alterations help distinguishing Alzheimer's and vascular dementia. *Front. Neurosci.* 12:274. doi: 10.3389/fnins.2018.00274
- Pellegrini, E., Ballerini, L., Hernandez, M. D. C. V., Chappell, F. M., González-Castro, V., Anblagan, D., et al. (2018). Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review. *Alzheimers Dement.* 10, 519–535. doi: 10.1016/j.dadm.2018.07.004
- Qureshi, M. N. I., Ryu, S., Song, J., Lee, K. H., and Lee, B. (2019). Evaluation of functional decline in Alzheimer's dementia using 3D deep learning and group ICA for rs-fMRI measurements. *Front. Aging Neurosci.* 11:8. doi: 10.3389/fnagi.2019.00008
- Reitz, C., Brayne, C., and Mayeux, R. (2011). Epidemiology of Alzheimer disease. *Nat. Rev. Neurol.* 7, 137–152. doi: 10.1038/nrneurol.2011.2
- Rousseeuw, P. J., and Hubert, M. (2011). Robust statistics for outlier detection. *WIREs Data Min. Knowl. Discov.* 1, 73–79. doi: 10.1002/widm.2
- Rubinov, M., and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 52, 1059–1069. doi: 10.1016/j.neuroimage.2009.10.003
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature* 323, 533–536.
- Rumelhart, D. E., and McClelland, J. (1987). "Learning internal representations by error propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations (MITP)*.
- Serrano-Pozo, A., Frosch, M. P., Masliah, E., and Hyman, B. T. (2011). Neuropathological alterations in Alzheimer disease. *Cold Spring Harb. Perspect. Med.* 1:a006189. doi: 10.1101/cshperspect.a006189
- Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V., and Greicius, M. D. (2012). Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cereb. Cortex* 22, 158–165. doi: 10.1093/cercor/bhr099
- Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., et al. (2006). Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. *Neuroimage* 31, 1487–1505. doi: 10.1016/j.neuroimage.2006.02.024
- Teipel, S. J., Grothe, M. J., Filippi, M., Fellgiebel, A., Dyrba, M., Frisoni, G. B., et al. (2014). Fractional anisotropy changes in Alzheimer's disease depend on the underlying fiber tract architecture: a multiparametric DTI study using joint independent component analysis. *J. Alzheimers Dis.* 41, 69–83. doi: 10.3233/JAD-131829
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978
- van Dijk, E. J., Prins, N. D., Vrooman, H. A., Hofman, A., Koudstaal, P. J., and Breteler, M. M. (2008). Progression of cerebral small vessel disease in relation to risk factors and cognitive consequences: Rotterdam Scan study. *Stroke* 39, 2712–2719. doi: 10.1161/STROKEAHA.107.513176
- van Dyck, C. H. (2018). Anti-amyloid- $\beta$  monoclonal antibodies for Alzheimer's disease: pitfalls and promise. *Biol. Psychiatry* 83, 311–319. doi: 10.1016/j.biopsych.2017.08.010
- Vinters, H. V., Zarow, C., Borys, E., Whitman, J. D., Tung, S., Ellis, W. G., et al. (2018). Review: vascular dementia: clinicopathologic and genetic considerations. *Neuropathol. Appl. Neurobiol.* 44, 247–266. doi: 10.1111/nan.12472
- Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst.* 1, 3–28. doi: 10.1016/0165-0114(78)90029-5
- Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D., and Alzheimer's Disease Neuroimaging Initiative (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* 55, 856–867. doi: 10.1016/j.neuroimage.2011.01.008
- Zheng, Y., Guo, H., Zhang, L., Wu, J., Li, Q., and Lv, F. (2019). Machine learning-based framework for differential diagnosis between vascular dementia and Alzheimer's disease using structural MRI features. *Front. Neurol.* 10:1097. doi: 10.3389/fneur.2019.01097

**Conflict of Interest:** The authors declare that this work was supported by the Italian Ministry of Health (RC2016) for data collection. CG received research funding from EPSRC, Horizon2020, Biogen Idec, Novartis, Wings for Life. These funders were not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication. Moreover, this study was also funded by the UK MS Society (programme grant number 984) and the ECTRIMS Postdoctoral Research Fellowship Programme to GC. The UK MS Society and ECTRIMS were not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

Copyright © 2020 Castellazzi, Cuzzoni, Cotta Ramusino, Martinelli, Denaro, Ricciardi, Vitali, Anzalone, Bernini, Palesi, Sinforiani, Costa, Miceli, D'Angelo, Magenes and Gandini Wheeler-Kingshott. 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.