European Journal of Nuclear Medicine and Molecular Imaging
https://doi.org/10.1007/s00259-018-4197-7

ORIGINAL ARTICLE

Accuracy and generalization capability of an automatic method for the detection of typical brain hypometabolism in prodromal Alzheimer disease

Fabrizio De Carli[1] · Flavio Nobili[2] · Marco Pagani[3,4] · Matteo Bauckneht[5,6] · Federico Massa[2] · Matteo Grazzini[2] · Cathrine Jonsson[4] · Enrico Peira[7] · Silvia Morbelli[5,6] · Dario Arnaldi[2] · for the Alzheimer's Disease Neuroimaging Initiative

Received: 7 June 2018 / Accepted: 16 October 2018

Abstract

Purpose The aim of this study was to verify the reliability and generalizability of an automatic tool for the detection of Alzheimer-related hypometabolic pattern based on a Support-Vector-Machine (SVM) model analyzing ^{18}F-fluorodeoxyglucose (FDG) PET data.

Methods The SVM model processed metabolic data from anatomical volumes of interest also considering interhemispheric asymmetries. It was trained on a homogeneous dataset from a memory clinic center and tested on an independent multicentric dataset drawn from the Alzheimer's Disease Neuroimaging Initiative. Subjects were included in the study and classified based on a diagnosis confirmed after an adequate follow-up time.

Results The accuracy of the discrimination between patients with Alzheimer Disease (AD), in either prodromal or dementia stage, and normal aging subjects was 95.8%, after cross-validation, in the training set. The accuracy of the same model in the testing set was 86.5%. The role of the two datasets was then reversed, and the accuracy was 89.8% in the multicentric training set and 88.0% in the monocentric testing set. The classification rate was also evaluated in different subgroups, including non-converter mild cognitive impairment (MCI) patients, subjects with MCI reverted to normal conditions and subjects with non-confirmed memory concern. The percent of pattern detections increased from 77% in early prodromal AD to 91% in AD dementia, while it was about 10% for healthy controls and non-AD patients.

Conclusions The present findings show a good level of reproducibility and generalizability of a model for detecting the hypometabolic pattern in AD and confirm the accuracy of FDG-PET in Alzheimer disease.

Keywords Alzheimer disease · MCI due to AD · FDG-PET · Discriminant analysis · Neuroimage classification · Classification and prediction · Neurodegenerative disorders · Support vector machine

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00259-018-4197-7) contains supplementary material, which is available to authorized users.

✉ Fabrizio De Carli
f.decarli@ibfm.cnr.it

[1] Institute of Molecular Bioimaging and Physiology, National Research Council, Largo Paolo Daneo, 3, 16132 Genoa, Italy

[2] Department of Neuroscience (DINOGMI), IRCCS Polyclinic San Martino-IST, University of Genoa, Genoa, Italy

[3] Institute of Cognitive Sciences and Technologies, CNR, Rome, Italy

[4] Medical Radiation Physics and Nuclear Medicine, Imaging and Physiology, Karolinska University Hospital, Stockholm, Sweden

[5] Department of Health Sciences (DISSAL), University of Genoa, Genoa, Italy

[6] Nuclear Medicine Unit, Polyclinic San Martino Hospital, Genoa, Italy

[7] National Institute of Nuclear Physics (INFN), Genoa, Italy

Published online: 31 October 2018

Introduction

The early diagnosis of Alzheimer Disease (AD) and the prognosis of patients suffering from Mild Cognitive Impairment (MCI) are a major challenge in clinical research for treatment strategies and in support to clinical trials for disease-modifying therapies. Among AD biomarkers, cerebral β-amyloidosis is considered a strong and early pathophysiological marker of the disease, preceding neurodegeneration and the cognitive symptoms, and possibly remaining relatively stable thereafter [1, 2]. Cerebral hypometabolism evaluated by 18-F-fluorodeoxyglucose (FDG)-PET is instead considered a marker of synaptic dysfunction associated with downstream neurodegeneration and in turn preceding brain structure damage (disclosed by Magnetic Resonance Imaging) and clinical symptoms [2]. In this context FDG-PET has been indicated as a major biomarker supporting the diagnosis of AD, assisting the prognosis of patients with MCI, whether at risk of developing AD Dementia (ADD), and monitoring disease severity [1, 3].

Statistical tools for image analysis based on the comparison between AD-related patterns and normal-control neuroimages support the role of FDG-PET in the diagnostic / prognostic process. The examination of single subject maps, expressed as statistical scores with respect to control groups, such as the ones provided by the SPM software (Statistical Parametric Mapping [4]) has been shown to increase the accuracy of both differential diagnosis and prognosis in neurodegenerative dementias [5]. The contribution of SPM-maps to the diagnostic process was significantly greater than simple visualization of raw FDG-uptake maps [6].

Based on statistical comparison between AD patients and healthy elderly subjects, different methods have been proposed for automatic classification of FDG-PET images aimed to support the diagnosis and to predict MCI to ADD conversion. Grounded on voxel-wise image analysis and the SPM approach, a global index of hypometabolism in AD-affected regions was proposed in 2002 [7] and is known as PMOD (PMOD Technologies) Alzheimer Discrimination Analysis tool (PALZ). PMOD software is commercially available, enabling the computation of PALZ, with a fixed threshold, to classify AD-related images, and has been applied to different multi-centric data with remarkable accuracy [8]. A similar approach, with more flexible method in defining and weighting AD-related hypometabolic clusters of voxels, was used to compute the Hypometabolic Convergence Index (HCI, [9]). Alternative methods analyze cerebral metabolism by a-priori grouping voxels into volumes of interest (VOIs) associated with homogeneous brain structures [3]. Our group analyzed FDG-PET data based on the partition of the whole brain into a set of meta-VOIs, starting from the VOIs defined by the Automated Anatomical Labeling (AAL) Atlas [10] and merging VOIs with similar anatomo-functional characteristics. We applied a Support Vector Machine (SVM) to discriminate patients with MCI due to AD (MCI-AD) from normal aging controls in a multi-centric European sample [11] and in a subsequent study on a local sample aimed at distinguishing MCI patients who later converted to ADD and MCI patients who did not convert within a follow-up time of at least 5 years [12]. Other approaches are reported in a review on the role of FDG-PET in the diagnosis of AD in MCI [13] which report a considerable variability in the resulting accuracy. The authors underline the need for further validation and standardization in order to apply FDG-PET analysis tools in clinical routine.

Another review examining neuroimaging based classification of AD by different modalities [14] suggested promising perspective for diagnostic and prognostic applications but underlined that generalizability and reproducibility of existing methods must be demonstrated. For this aim, the growth and standardization of extensive archives of neuroimages play an important role. The Alzheimer's Disease Neuroimaging Initiative (ADNI, http://adni.loni.usc.edu/) provide a large dataset of clinical and neuroimage data from normal aging and AD patients at different stages of the disease and with extensive follow-up. Analogous datasets have been recorded in Europe by the European Alzheimer's Disease Consortium (http://www.eadc.info/sito/pagine/home.php) [15] and by the Network for Standardization of Dementia Diagnosis (NEST-DD; [16]), and other initiatives are in progress. The increasing homogeneity of acquisition protocols and a remarkable level of test-retest reliability, also with respect to cognitive and neuropsychological measure [17], enable further developing and validation of FDG-PET based diagnostic tools. In this context, a comparison between different FDG-PET-based automatic methods for AD classification was performed using different datasets and suggested high potential accuracy based on the analysis of Receiver Operating Characteristic (ROC) curves [18].

The aim of this study was to verify the reliability and generalizability of the automatic tool for AD diagnosis and MCI prognosis based on SVM analysis of meta-VOIs. For this purpose, the tool we developed by analyzing a relatively homogeneous local sample of controls and patients with MCI-AD and ADD [12], was applied to a larger and multicentric sample drawn from the ADNI archive and including healthy elderly subjects, MCI-AD patients at different stages and ADD patients.

Material and methods

Participants

Genoa dataset

The automatic tool applied in this study is an SVM classifier processing FDG-uptake values as computed for an extensive

set of brain volumes of interest. It was previously developed [12, 19] and in this study was specifically fitted to discriminate MCI-AD and ADD patients from healthy controls in a sample largely overlapping the ones analyzed in previous studies [12, 20]. The sample included 42 healthy elderly subjects (NA, Normal Aging), 95 MCI-AD patients, who had a baseline FDG-PET during the MCI stage and converted to ADD during the follow-up period, and 55 patients with mild ADD diagnosis at baseline PET. MCI-AD patients were further divided into early MCI (eMCI, 37 patients), who converted to ADD later than 2 years since baseline FDG-PET and late MCI (lMCI, 58 patients) who converted within the first 2 years. A further group of 27 MCI patients, who had not converted to ADD within 5 years after the baseline scan (ncMCI), were not included in the training set due to its heterogeneity (presence of very-late converters) but were considered in the testing phase. General information regarding subject groups are reported in Table 1a.

MCI and ADD patients were subjects referred to the University neurology memory clinic of IRCCS Polyclinic San Martino Hospital (Genoa – Italy) for a first assessment of cognitive complaints. All patients underwent an extensive diagnostic work-up, including a thorough clinical examination, the Mini-Mental State Examination (MMSE) for global cognition, and a battery of neuropsychological tests, as well as morphological (MRI) and functional (FDG-PET) neuroimaging. Patients with memory impairment confirmed by a memory test but not demented were included in the MCI group according to the Petersen's MCI criteria [21, 22]. Patients with significant impairment in daily activities, as assessed by questionnaires for daily living (ADL), instrumental ADL (IADL), and by the Clinical Dementia Rating (CDR) scale, but with mild dementia (MMSE >18), were included in the AD group, according to NIA-AA criteria [23]. The clinical follow-up was set at least every 6 months, including clinical examination and formal interviews for ADL, IADL, and CDR, and MMSE administration. Detailed neuropsychological tests were checked during the follow-up according to the clinical needs.

The control group was composed of voluntary healthy subjects approximating the patient groups as for the distribution of age, gender and education level. They were carefully examined for medical condition and medical history and underwent the same neuropsychological battery as patients, FDG-PET, and MRI neuroimaging. Only subjects with MMSE score greater than 26 and CDR of zero were included. Clinical condition of healthy volunteers was monitored during follow-up period by clinical examination and formal interviews at least yearly.

Exclusion criteria included major psychiatric or neurological disorders, uncontrolled arterial hypertension, diabetes, anaemia and malignancy and are detailed in a previous paper [20] along with extensive information on the sample.

The institutional review board of the University of Genoa approved the recording and data treatment procedures, and all subjects gave written informed consent to undergo FDG-PET in the framework of a long-term observational study, in accordance with the Declaration of Helsinki.

This local set of data is hereinafter referred to as Genoa-set.

Table 1 General information on the subjects in each group and in the two datasets

Group	Size	Males	%	Age mean ± std	MMSE mean ± std
A – Genoa-set					
NA	42	11	26.2	68.3 ± 9.6	29.2 ± 0.9
ncMCI	27	15	55.6	71.9 ± 6.5	26.9 ± 1.4
eMCI	37	10	27.0	74.6 ± 6.9	26.3 ± 1.7
lMCI	58	21	36.2	75.6 ± 6.5	25.8 ± 1.9
ADD	55	20	36.4	73.3 ± 7.3	19.2 ± 4.1
B – ADNI-set					
NA	150	65	43.3	75.3 ± 6.7	29.0 ± 1.4
SMC	66	24	36.4	72.1 ± 5.3	28.9 ± 0.9
recovMCI	33	15	45.5	68.7 ± 8.4	29.0 ± 1.3
ncMCI	87	53	60.9	70.8 ± 7.4	28.0 ± 2.2
eMCI	22	12	54.5	73.3 ± 6.2	26.7 ± 1.6
lMCI	76	40	52.6	74.4 ± 7.4	21.3 ± 5.5
ADD	99	58	58.6	74.2 ± 8.1	18.3 ± 4.1

eMCI and lMCI groups include only patients who converted to ADD during the follow-up. SMC group include only patients who had been diagnosed as normal in all visits

Abbreviations: *ADNI-set* sample of data drawn from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, *Genoa-set* sample recorded at the University neurology memory clinic of IRCCS Polyclinic San Martino Hospital (Genoa – Italy), *MMSE* Mini-Mental State Examination

Group labels: *NA* Normal Aging, *SMC* Significant Memory Concern, *recovMCI* recovered MCI, i.e. patients with an initial diagnosis for mild cognitive impairment (MCI) later on revised to normal, *ncMCI* not converted MCI patients, MCI confirmed without dementia after 4–5 years of follow-up, *eMCI* MCI patients in the early stage, *lMCI* MCI patients in advanced stage, *ADD* patients with Alzheimer Disease Dementia

ADNI data set

The testing set (ADNI-set) was drawn from the ADNI archive by selecting FDG-PET images included in the stage ADNI2 of the ADNI project for homogeneity of associated information. It was made up of 533 subjects including 150 cognitively normal, 66 with Significant Memory Concerns (SMC), 218 MCI at different stages and 99 ADD.

This classification and following subgrouping were drawn from information stored in the ADNI database and in particular:

- a basic classification was assumed from the advanced search tool for ADNI archive where the following classes

were defined (in http://adni.loni.usc.edu/study-design/background-rationale/)

- **NA**: cognitively normal subjects without signs of cognitive impairment or depression;
- **SMC**: with subjective memory complaints not confirmed by examination (CDR = 0) nor by informants.
- **eMCI**: Early MCI, with light sign of memory impairment as evaluated using the Wechsler Memory Scale Logical Memory II;
- **lMCI**: Late MCI, more severe memory impairment without signs of dementia;
- **ADD**: meeting the NINCDS/ADRDA criteria for probable Alzheimer dementia.

- The classification was then refined by processing diagnostic information associated with the sequence of follow-up control visits of each subject. Subjects in the NA and ADD groups were included if the diagnosis was confirmed in all available visits while SMC subjects were included if they had been diagnosed as normal in all visits. MCI patients were included in the study if they fell into one of the following categories:

- *recovered MCI* (**recovMCI**) if the initial MCI diagnosis was changed to normal (NA);
- *non-converted MCI* (**ncMCI**) if the initial MCI diagnosis was confirmed after more than 4 years since FDG-PET acquisition;
- *Early or Late converted MCI* **(eMCI or lMCI)** if their diagnosis was changed to ADD as confirmed at the last available visit.

General information about subjects in the different groups is reported in Table 1b.

FDG-PET data acquisition and preprocessing

FDG-PET images in the Genoa-set had been acquired according to the guidelines of the European Association of Nuclear Medicine [24]. Approximately 45 min after the injection of about 200 MBq of 18F-FDG, the acquisition was started by a SIEMENS Biograph 16 PET/CT equipment and lasted 15 min in 3-D mode. An ordered-subsets expectation-maximization algorithm was applied to reconstruct images with a voxel size of 1.33 × 1.33 × 2.00 mm, followed by attenuation correction based on CT scan. Spatial normalization into MNI space (from the Montreal Neurological Institute) was then performed by affine and nonlinear processing using SPM12 software (Wellcome Department of Cognitive Neurology, London, UK) with an FDG PET template optimized for dementia patients [25]. Spatial normalization was followed by 8-mm smoothing with isotropic Gaussian filter. Further details on image reconstruction can be found in [20].

From the ADNI database, we downloaded the preprocessed images associated with the selected subjects. Preprocessing included dynamic co-registration of images acquired in consecutive time frames, averaging, reorientation along the anterior - posterior commissure (AC-PC line) and filtering with a scanner-specific filter function to produce images of a uniform isotropic resolution of 8 mm FWHM.

ADNI images were downloaded in Analyze format and then processed by SPM12 software for spatial normalization into the same MNI space used for the Genoa-set.

Activity distribution in regions of interest

For each image, from both Genoa-set and ADNI-set, mean FDG uptake values were computed for a set of 90 anatomical volumes of interest (VOIs, 45 homologue regions for each hemisphere) as defined by the AAL Atlas [10]. Regional values were normalized to the average activity of cerebellum, as this is considered a mainly spared region in Alzheimer physiopathology [26, 27]. The number of variables was further reduced by merging AAL volumes with similar anatomo-functional characteristics into a set of 29 meta-VOIs (14 in each hemisphere plus the Vermis, as listed in Table 2) with the aim to favor emergence of the brain structures most involved in the pathological process.

Data analysis and statistics

A descriptive statistic was used to summarize sample composition as for age, gender and MMSE score; t-test, one-way analysis of variance (ANOVA) and chi-square test were applied to explore relevant differences. The differences in mean uptake values between each patient group and the NA control group were explored in each dataset using the voxel-based 2-sample t-test implemented in SPM12. Clusters of regional differences were identified with a significance threshold set at 0.05, corrected for multiple comparisons by family-wise error (FWE) option, and with minimum cluster size of 100 voxels. Age and gender were considered as covariates.

A preliminary General Linear Model (GLM) for repeated measurements was applied to meta-VOI values in each dataset in order to estimate the effect of age and gender on the distribution of FDG uptake and consequently adjust the data for each subject. The preliminary GLM analysis was restricted to the control group in order to estimate the physiologic effect independently from the prevalence of the pathology, which increases with age and is different between genders. The estimated regression coefficients were then applied to all subjects in the respective dataset and age/gender corrected data were used in the following analysis.

Table 2 List of the 29 meta-VOIs (14 in each hemisphere plus the Vermis) used as input to the classification model

DEV-fit	ADNI-fit	meta-VOI	Including
R L A		Occipital Cortex	Calcarine/ Lingual/ Inferior Occipital/ Middle Occipital/ Superior Occipital Gyri
R L		Basal ganglia	Putamen/ Pallidum/ Caudate
R L A	R L	Amygdala/Insula/ Hippocampus	Parahippocampal gyrus
R	R L A	Orbito-frontal Cortex	Inferior Frontal/ Medial Frontal/ Middle Frontal Gyri
R	A	Frontal Cortex	Middle Frontal/ Superior Frontal/ Superior-Medial Frontal/ Superior-Orbital Frontal/ Inferior Frontal Gyri
L	A	Cuneus/ Fusiform Gyrus/ Precuneus	
R L	R A	Pre/post-central cortex	Postcentral Gyrus/ Precentral Gyrus/ Supplementary Motor Area
	L	Parietal Lobe	Inferior Parietal/ Superior Parietal Gyri
L		Anterior Cingulate Gyrus	
R L	R L A	Posterior Cingulate Gyrus	
R L A	L A	Temporal Lobe	Inferior Temporal/ Middle Temporal/ Superior Temporal Gyri
	R	Temporal Pole	Middle Temporal Pole/ Superior Temporal Pole Gyri
	L	Thalamus	
		Cerebellum	
		Vermis	

The meta-VOIs selected in the training phase by fitting on the Genoa-set and ADNI-set respectively are marked on the first two columns with R: left hemisphere, R: right hemisphere, A: interhemispheric asymmetry

In order to discriminate patients with probable Alzheimer Disease from control subjects (cognitively normal or non-declining MCI), we applied an SVM model with linear kernel. The model was trained on Genoa-set by contrasting cognitively normal controls with AD patients (ADD and MCI-AD). The predictors in the model were FDG uptake values in the meta-VOIs; considering that asymmetric patterns of resting-state brain metabolism are often indicative of pathological processes [28, 29], we also considered asymmetries as potential predictors:

$$asy_{voi} = \frac{abs\left(x_{voi,right} - x_{voi,left}\right)}{\left(x_{voi,right} + x_{voi,left}\right)}$$

where $x_{voi,right\ /\ left}$ represents the normalized uptake value for a meta-VOI in the right or left hemisphere and abs is the absolute value function.

The performance of the SVM model on the Genoa-set was evaluated using 25-fold cross-validation and a step-wise procedure was applied to select the best set of predictors among the ones associated to the 29 meta-VOIs and 14 asymmetries. At each step the performance of the model was estimated by the Youden index (= sensitivity + specificity −1). The global performance of the model was described by ROC curve analysis and relevant parameters were estimated with their 95% confidence intervals (CIs) as evaluated using specific methods: the Wald interval for sensitivity, specificity and accuracy [30], the Simel method [31] for positive and negative likelihood ratio, the Chen approach [32] for Youden Index and the bootstrap method described by Qin and Hotilovac [33] for the Area Under ROC Curve (ROC-AUC).

The SVM model fitted on the Genoa-set was then applied to the ADNI-set in order to verify its generalization capabilities. The distribution of positive and negative subjects, as classified by the SVM model, was evaluated for each ADNI group and an estimation of global sensitivity, specificity and accuracy was performed considering cognitively normal subjects (including NA and SMC) as opposed to MCI-AD (including eMCI and lMCI) and ADD.

For a further evaluation of the model and its potential, the role of the two datasets was reversed and the model was fitted to the ADNI-set, again contrasting cognitively normal subjects with MCI-AD and ADD. Accuracy parameters were computed, and the ROC-AUCs evaluated for the two datasets were compared by computing the z-value according to the Hanley and McNeil method [34]. The ADNI-fitted model was then applied to the data in the Genoa-set. As in a previous case, the distribution of positive and negative subjects in each group was evaluated along with a global estimation of sensitivity, specificity and accuracy.

The main steps covered in data processing and statistical analysis are summarized in Fig. 1.

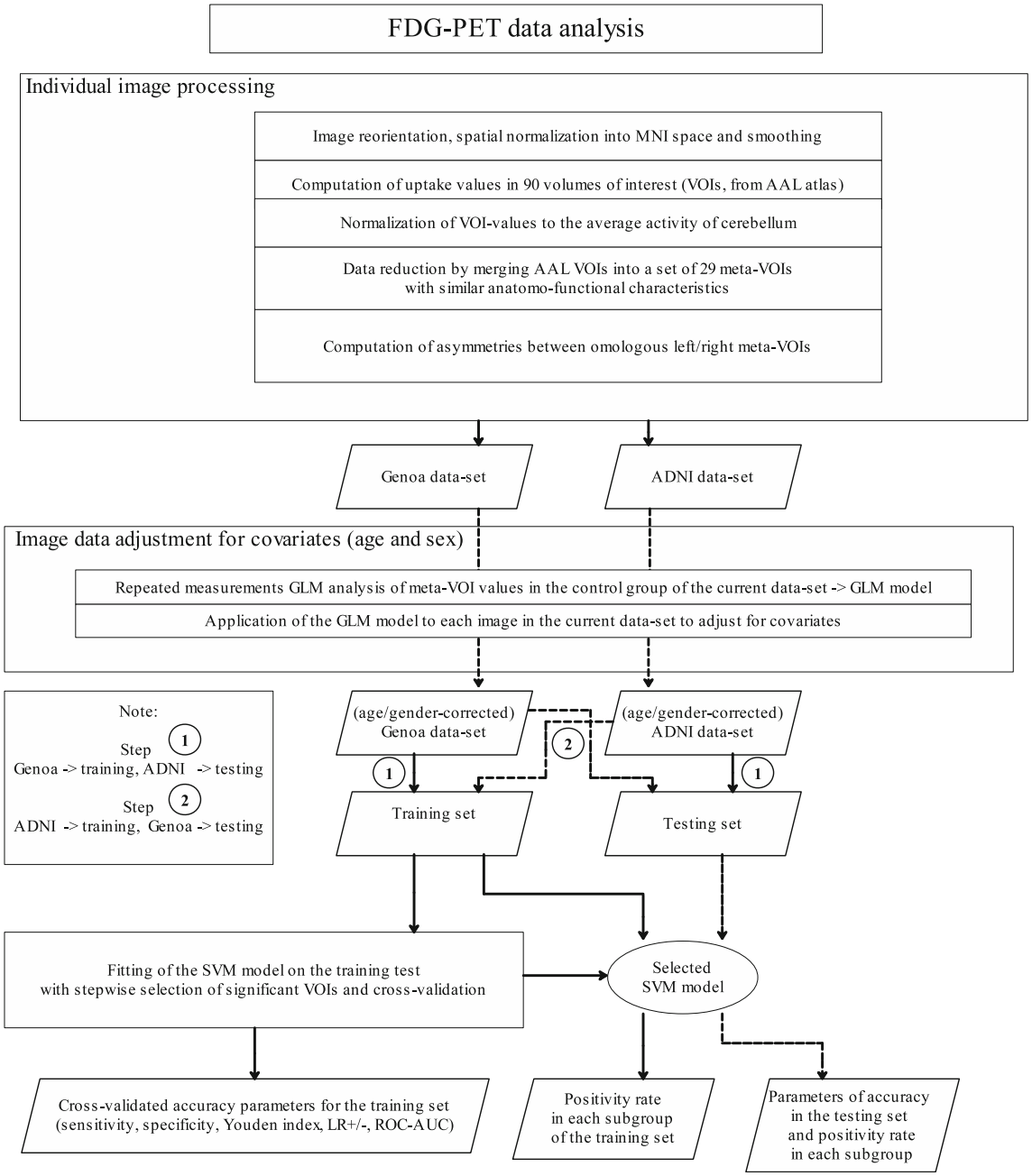

Fig. 1 Flowchart summarizing FDG-PET data processing and statistical analysis

The binary classification performed by the SVM model was based on the application of a zero-threshold to a global score, which was computed for each image: the distribution of scores among groups has been examined by box-plot graphs.

In order to explore the relationship between SVM based classification of FDG-PET data and underlying presence of amyloid beta protein, mean cortical florbetapir uptake was drawn from the ADNI dataset, subjects were categorized as normal or abnormal following Landau et al. [35] and cross-classification (SVM-based versus amyloid-based) was performed for each subject group.

Data analysis was performed using Matlab 2017b (MathWorks, Natick, MA) and its Statistics and Machine Learning Toolbox.

Results

Population characteristics

General information about subject distribution in the two datasets, age, gender and MMSE score, is reported in

Table 1. Analogous groups in the two datasets have been coupled, for descriptive purposes, in spite of some differences in their definition: the difference between early and late MCI-AD is based on time to conversion in the Genoa-set but on severity of impairment in the ADNI-set. The selection of ncMCI is slightly different in the two datasets due to different follow-up time. However, the partition of the sample has strong similarities. The between group distribution of age and gender is fairly variable in both datasets (between group ANOVA for age: Genoa-set: $F_{4,214} = 6.44$ $p < 0.0001$; ADNI-set: $F_{6,526} = 6.88$, $p < 0.0001$; between group chi-square for gender: Genoa-set: $\chi^2_4 = 7.55$ $p = 0.11$,N.S.; ADNI-set: $\chi^2_4 = 15.31$, $p < 0.02$). AS for the comparison of mean age between the two datasets within each group, the difference was not significant for all groups but NA: control subjects in the Genoa-set were younger than in the ADNI-set ($t = -5.450$, $p < 0.0001$). The proportion of males was higher in the ADNI-set (with $p < 0.05$ for NA, eMCI and ADD groups). MMSE was obviously different among different groups within each dataset, while it was similar between datasets within each group except for ncMCI, who were more impaired in Genoa-set ($t = -2.37$ $p < 0.025$) and for lMCI, who were quite more impaired in the ADNI-set ($t = 5.7$, $p < 0.0001$).

FDG-PET mean values

Voxel-based comparison of mean FDG-PET values showed clusters of hypometabolism, as compared to NA, in MCI-AD patients (both eMCI and lMCI) and in ADD, while no significant clusters of hypometabolism were found in ncMCI nor in recovMCI and SMC subjects. The main areas of hypometabolism, involving cingulate gyrus and temporo-parietal lobes, are shown in Fig. 2.

GLM analysis, as applied to NA group, showed a significant age*region interaction (Genoa-set: $F_{28,1120} = 2.64$, $p < 0.02$; ADNI-set: $F_{28,6412} = 9.56$, $p < 0.0001$) but no gender effect on normalized FDG-PET values in both datasets. All data were consequently corrected for age effect and corrected data were used in the following analysis.

Support-vector-machine classification

The SVM model was first trained on the Genoa-set by contrasting NA group with all AD (MCI-AD and ADD), applying a stepwise selection of meta-VOIs (or relevant asymmetries) and measuring the discrimination performance with cross-validation. A set of 19 variables (16 meta-VOIs and three asymmetries, as marked in Table 2) and supplementary Figure 1 was selected, producing a high discrimination between the two groups, with 96.0% sensitivity, 95.2% specificity (Youden index = 0.912, ROC-AUC = 0.979; Fig. 3). Table 3a and b report detailed classification parameters including the percentage of positive and negative cases in each subgroup of the Genoa-set. When this model was applied to the ADNI-set, we obtained 83.2% sensitivity and 90.0% specificity (Youden index = 0.738). Table 3c reports the percentage of positive and negative cases in each subgroup of the ADNI-set.

When SVM model was trained on ADNI-set, 17 variables were selected (11 meta-VOIS and six asymmetries, as marked in Table 2 and supplementary Figure 1) resulting in 87.3% sensitivity, 92.1% specificity (Youden index = .794, ROC-AUC = 0.94). The difference between this AUC and the one relevant to the Genoa-set was statistically significant ($z = 2.52$, $p < 0.01$). Table 4a and b report detailed classification parameters including the percentage of positive and negative cases in each subgroup of the ADNI-set. When this model was applied to the Genoa-set, we obtained 88.1% sensitivity, 88.0% specificity (Youden index = 0.761). Table 4c reports the percentage of positive and negative cases in each subgroup of the Genoa-set.

The distribution of SVM scores, following the training on the ADNI-set and Genoa-set, in the different groups of subjects, is reported in Fig. 4.

Explorative analysis of amyloid data

Amyloid deposition data were available for the ADNI dataset and, according to [35], the rate of florbetapir-positive subjects was 30.7% in the NA group, 86.9% in ADD and 87.8 among MCI converters (either early or late; further details in supplementary Table 1). Among the 14 SVM-positive subjects in the NA group (9.3%), nine were florbetapir-positive while among the nine SVM-negative subjects in the ADD group (9.1%), five were florbetapir-negative. Among the 16 SVM-negative MCI converters only two were florbetapir-negative.

Discussion

Cerebral hypometabolism, as evaluated by FDG-PET, is considered a major biomarker which could support complex AD diagnosis and prediction of the course of the disease [1, 5]. FDG-PET can be also performed for patient selection and monitoring during clinical trials for disease-modifying therapies. Of note, it can also assist in discriminating among all main forms of dementing conditions, not only between AD and dementias without brain amyloidosis (but dementia with Lewy bodies), as it happens for amyloid PET tracers. For reliability and effectiveness of FDG-PET application, the methods for image analysis and classification have to be verified for reproducibility and generalizability in extended datasets with heterogeneous origin [14].

The objective of the present study was to verify the generalization capability of the method we developed to discriminate MCI-AD patients from normal aging or ncMCI subjects. The application of the model, previously fitted on a selected

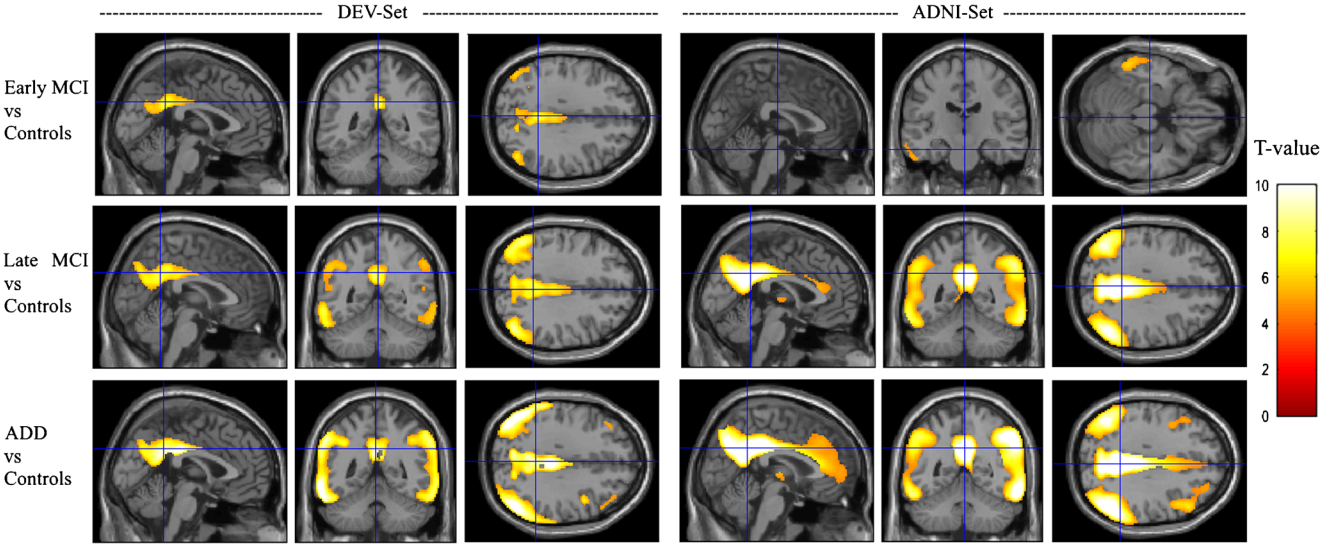

Fig. 2 Topographic representation of clusters in which mean 18-F-FDG up-take was significantly lower in a patient group with respect to the proper control group. Significant clusters were detected by voxel-based 2-sample t-test in SPM12, with significance threshold set at 0.05, corrected for multiple comparisons with familywise error option and a minimal cluster size of 100 voxels. Clusters were superimposed on the Montreal Neurologic Institute template in coronal (left), sagittal (middle), and transversal (right) views. The comparisons were performed within each dataset: the Genoa-set (recorded at the Genoa University neurology memory clinic) on the left and the ADNI-set (drawn from the Alzheimer's Disease Neuroimaging Initiative database) on the right. Significant differences were found for patients with prodromal AD (eMCI and lMCI groups) and for patients with Alzheimer Disease Dementia (ADD group). No clusters of significant differences were found for the other patient groups (SMC, recovMCI and ncMCI) with respect to normal aging controls

monocentric sample (Genoa-set) to an independent sample from ADNI archive (ADNI-set) showed a good accuracy (83.2% sensitivity, 90.0% specificity) which is however lower than the one we obtained in the original sample (96.0% sensitivity, 95.2% specificity). Furthermore, when the model was fitted to the ADNI-set the accuracy of classification in this dataset increased (87.3% sensitivity, 92.1% specificity) and maintained a high accuracy when applied to the Genoa-set (88.1% sensitivity, 88.0% specificity). It's worth noting that for both datasets the fitting procedure was associated with cross-validation. Such technique favors a more reliable estimation of accuracy in that it enables the separation between the training and testing stage of the method, also in case of relatively limited sample size. In this way, accuracy estimation is nearly independent from the characteristics of the particular sample (the risk of overfitting is reduced) and reflects the characteristics of the reference population. However, in case of Genoa-set the reference population is more homogeneous

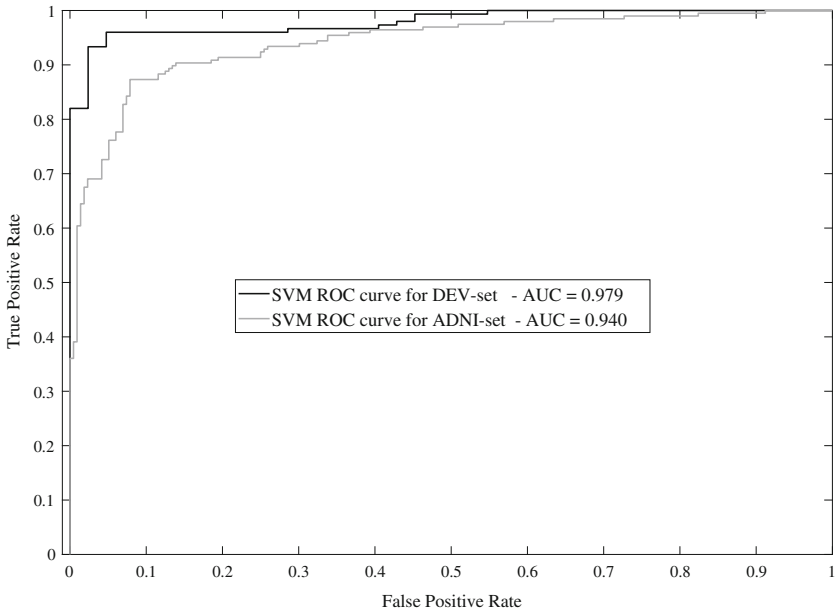

Fig. 3 Receiver-operating-characteristic curves obtained by SVM classifier as applied to the two different datasets: the Genoa-set (recorded at the Genoa University neurology memory clinic) and the ADNI-set (drawn from the Alzheimer's Disease Neuroimaging Initiative database). The difference between the two Areas Under the Curve (AUC) was statistically significant ($z = 2.52$, $p < 0.01$)

Table 3 Classification of subjects based on the original SVM model trained on the dataset recorded at the Genoa University memory clinic

A - SVM cross-validation (Genoa-set)			B – SVM application (Genoa-set)				C – SVM application (ADNI-set)			
	Estimate	Conf.Int.	Group	Size	% SVM+	% SVM-	Group	Size	% SVM+	% SVM-
Sensitivity	.960	.929–.991	NA	42	2.38	97.62	NA	150	10.00	90.00
Specificy	.952	.888–1.00					SMC	66	12.12	87.88
Accuracy	.958	.930–.987					recovMCI	33	6.06	93.94
Youden i.	.912	.844–.981	ncMCI	27	29.63	70.37	ncMCI	87	17.24	82.76
LR+	20.16	5.12–77.9	eMCI	37	89.19	10.81	eMCI	22	72.73	27.27
LR-	.042	.019–.092	lMCI	58	89.66	10.34	lMCI	76	84.21	15.79
Roc-AUC	.979	.844–.992	ADD	55	92.73	7.27	ADD	99	83.84	16.16

A: accuracy parameters estimated after cross-validation in the training set; B percent of positive and negative tests for each subject group in the training set after a global fitting; C: percent of positive and negative tests for each subject group in the testing set following the global fitting in the training set.

Abbreviations: *LR+* positive likelihood ratio, *LR-* negative likelihood ratio

Group labels: *NA* Normal Aging, *SMC* Significant Memory Concern, *recovMCI* recovered MCI, i.e., patients with an initial diagnosis for mild cognitive impairment (MCI) later on revised to normal, *ncMCI* not converted MCI patients, MCI confirmed without dementia after 4–5 years of follow-up, *eMCI* MCI patients in the early stage, *lMCI* MCI patients in advanced stage, *ADD* patients with Alzheimer Disease Dementia

in that it is made up of neuroimages acquired with the same equipment, from subjects examined by the same medical team, while ADNI-set is representative of data from multiple centers, with variable equipment and clinical teams. The generalization capability is clearly dependent on the level of homogeneity and standardization of protocols for data acquisition and processing. From this point of view the finding that a model fitted on one dataset maintains a fair good accuracy when applied to the other one is promising for further developments.

In a review on the potential role of FDG-PET in the early diagnosis of AD [13], the authors included 14 studies using visual image evaluation possibly supported by quantitative and statistical analysis and estimating accuracy by cross-validation. The review reports a remarkable variability in the detection of AD in MCI patients and, based on a globally fitted ROC-curve, estimated a sensitivity of 76% at 82% specificity. The authors observed a fair homogeneity in the protocol followed by different studies but heterogeneity in the interpretation of data and stressed the lack of an a-priori threshold for the analysis of quantitative data. The present study fulfills the need to estimate the model accuracy by a-priori fixing the classification threshold in that the model is fully defined by the fitting on a dataset and then tested on another completely independent dataset, and the results are in the highest range of the accuracy reported in the cited review.

A comparative study on different datasets [18] examined three automatic methods, i.e.: PALZ [7], HCI [9] and

Table 4 Classification of subjects based on the SVM model trained on the dataset drawn from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database

A - SVM cross-validation (ADNI-set)			B – SVM application (ADNI-set)				C – SVM application (Genoa-set)			
	Estimate	Conf.Int.	Group	Size	% SVM+	% SVM-	Group	Size	% SVM+	% SVM-
Sensitivity	.873	.827–.920	NA	150	9.33	90.67	NA	42	11.90	88.10
Specificy	.921	.885–957	SMC	66	3.03	96.67				
Accuracy	.898	.869–.927	recovMCI	33	6.06	93.94				
Youden i.	.794	.741–.848	ncMCI	87	21.84	78.16	ncMCI	27	44.44	55.56
LR+	11.09	7.01–17.6	eMCI	22	77.27	22.73	eMCI	37	78.38	21.62
LR-	0.137	.095–.199	lMCI	76	85.53	14.47	lMCI	58	91.38	8.62
Roc-AUC	.940	.909–.959	ADD	99	90.91	9.09	ADD	55	90.91	9.09

A: accuracy parameters estimated after cross-validation in the training set; B percent of positive and negative tests for each subject group in the training set after a global fitting; C: percent of positive and negative tests for each subject group in the testing set following the global fitting in the training set.

Abbreviations: *LR+* positive likelihood ratio, *LR-* negative likelihood ratio

Group labels: *NA* Normal Aging, *SMC* Significant Memory Concern, *recovMCI* recovered MCI, i.e., patients with an initial diagnosis for mild cognitive impairment (MCI) later on revised to normal, *ncMCI* not converted MCI patients, MCI confirmed without dementia after 4–5 years of follow-up, *eMCI* MCI patients in the early stage, *lMCI* MCI patients in advanced stage; ADD: patients with Alzheimer Disease Dementia

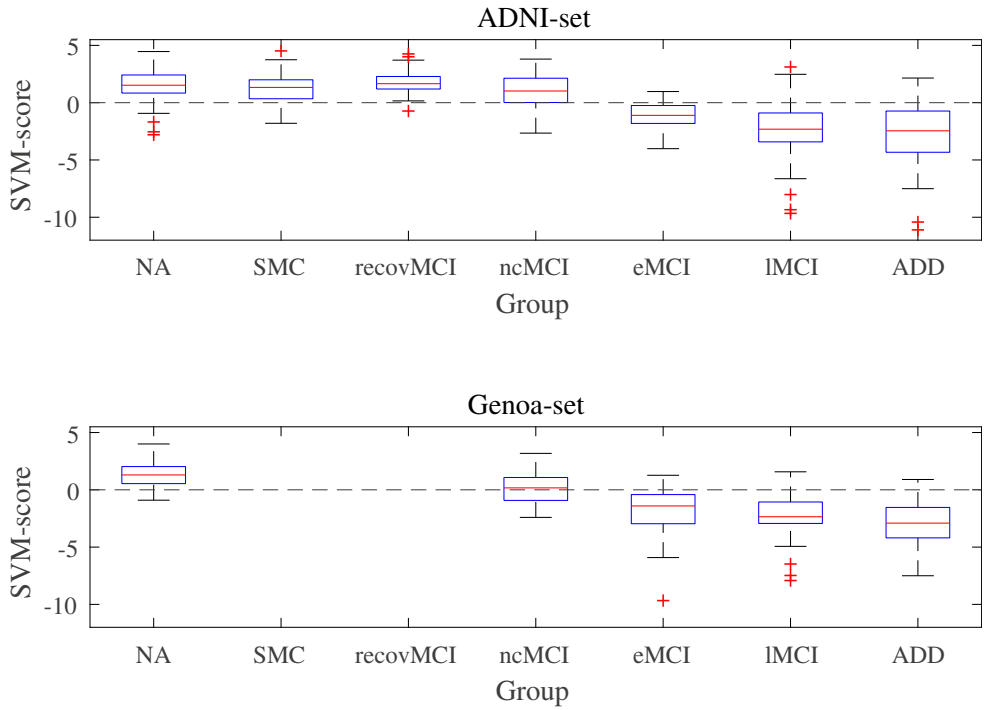

Fig. 4 Boxplot depicting the distribution of SVM scores for each subject group in the two datasets. For each column we can find the median (line inside the box), the interval between 25° and 75° percentiles (lower and upper side of the box) and the extension to extreme values (whiskers) and possible outliers('+'). *Abbreviations*: ADNI-set: sample of data drawn from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Genoa-set: sample recorded at the Genoa University neurology memory clinic *Group labels*: NA: Normal Aging; SMC: Significant Memory Concern; recovMCI: recovered MCI, i.e., patients with an initial diagnosis for mild cognitive impairment (MCI) later on revised to normal; ncMCI: not converted MCI patients, MCI confirmed without dementia after 4–5 years of follow-up; eMCI: MCI patients in the early stage; lMCI: MCI patients in advanced stage; ADD: patients with Alzheimer Disease Dementia

metaROI [3]. The authors reported high accuracy in discriminating MCI-AD and mild ADD patients from a normal-aging control group: the estimated ROC-AUC varied from 0.77 to 0.98 with variable accuracy and different ranking of the three indexes as function of disease stage and of the considered dataset. The results of this study are promising for the three indexes in spite of two methodological limitations. Firstly, all accuracy estimations were performed by independent ROC-curve analysis for each disease-stage subgroup in each dataset, without establishing a common threshold and, seemingly, without cross-validation. Secondly, all patient subgroups were compared with a unique selected control group, without an independent estimation of specificity based on normal aging subjects in the different datasets. These limitations can be overcome by further studies with head-to-head comparison of different classification methods in multicentric datasets in order to establish generalizability of FDG-PET-based automatic models.

The PALZ index, which is available as commercial software with a predefined threshold for AD pattern detection, has also been applied in other studies. Haense et al. [8] applied PALZ to discriminate ADD patients from healthy controls in a sample from the ADNI dataset, obtaining 83% sensitivity and 78% specificity, and in a sample from the NEST-DD dataset, with a 78% sensitivity and 94% specificity. These findings confirm the fair accuracy of this index but also its decrease and variability when applied to a new dataset, independent from the training sample.

The results of the present study suggest that the training on a more extended and less homogeneous dataset (the ADNI-set) could yield a somewhat lower accuracy on the (cross-validated) training set, but it is associated with a lower decrease when applied to an independent testing set (the Genoa-set). The training of a promising model on an extensive dataset, collected from different centers but with strictly standardized protocols, could yield to accurate and generally-applicable automatic tools.

We discussed so far the accuracy parameters derived from discriminating AD-patients from controls (AD-test). However, the rate of subjects who are positive or negative at the AD-test has been presented for all subgroups drawn from the two datasets. A remarkable difference was found between eMCI and lMCI in both datasets, in spite of the different definition used, as it was based on time-to-conversion in the Genoa-set and on cognitive performance in the ADNI-set. In the testing stage, lMCI patients were classified as AD-positive at a rate (84–91%) close to ADD patients while the rate of AD-positive among eMCI was lower (73–78%). The higher positivity in lMCI is in agreement with previous studies suggesting an

hypometabolic pattern of lMCI closely approaching ADD [9, 36, 37] while a reduced pattern of hypometabolism can be find in eMCI [13, 38]. The progressive worsening of metabolism through the three groups of AD patients (eMCI, lMCI and ADD) was highlighted by the decreasing trend of the score drawn from the SVM model (Fig. 4). The automatic models might be refined to fit better the early stage of the disease, also enlarging the eMCI-AD group, which is underrepresented in the present datasets. The analysis of a dataset with a larger representation of eMCI-AD patients may also enable fitting the model to the subgroups, focusing on patterns characteristic of the early phase of the disease to increase model sensitivity (see supplementary figure 2 and 3 for an explorative description of different trends). Another group is made up of MCI patients who did not convert to ADD within a long follow-up time (at least 4 years). The variable classification of these patients probably reflects the coexistence of non-AD patients with very late converters. On the other side, there are two other groups in the ADNI-set: SMC, who complained of memory impairment but were scored within normal range for cognition, and recovMCI, who were initially diagnosed as MCI, but later revised to normal. It is worth noting that in both cases they were negative to the AD-test at around 90% or even more (Table 3). Subjective memory complaints are considered a risk factor for future memory impairment and Alzheimer disease [39], however, only subjects confirmed as cognitive normal after a mean follow-up time of 1.8 years were included in the SMC group and amyloid data indicated possible amyloidosis for 35.9% of these patients, only slightly more than the NA group (31.3%) and recovMCI (27.3%). Therefore, we found a similar trend in these three groups, suggesting a limited probability of developing AD symptoms in the near future, the higher risk probably concerning subject positive on both AD-test and amyloidosis.

The training stage of the SVM model envisaged the selection of a set of meta-VOIs suitable to reach high accuracy. The best results were reached with a high number of meta-VOIs including regions which are known to be hypometabolic in AD, such as temporo-parietal lobes, amygdala-hippocampus-insula and posterior cingulate gyrus, but also other regions, such as the pre- post-central cortex, which are generally spared in AD (Table 2 and supplementary figure 1). Selected regions also extended beyond the clusters where significant differences were found at voxel-based comparison of mean values, as this clusters involved a delimited set of regions in the expected brain structures (Fig. 2). Our VOI-based model was probably sensitive to secondary hypometabolic areas and, moreover, also spared regions that played a role in the building of the statistical model as the high difference between affected and spared regions can be a strong index of hypometabolism [12, 19]. On the same basis, the model included asymmetry values, which often play an important role also in the visual evaluation of abnormal metabolic patterns, particularly in the early stages of the disease [28, 29].

The inclusion in the model of data from asymmetries and spared regions is relevant also considering the problem of data normalization. FDG-PET regional values are normalized to reduce the effect of technical factors and inter-subject variability, looking for reference values minimally affected by pathological processes [40]. In this study, the average activity of the cerebellum was preferred to the cerebral global mean which can be significantly reduced for the contribution of affected regions [27]. Some preliminary tests on data used in the present study showed slightly higher sensitivity in discriminating AD from NA when cerebellar normalization was used with respect to global mean normalization. However, cerebellar metabolism may in turn be mildly affected by the presence of amyloid plaques [41, 42] and by cross diaschisis [43] and the inclusion in the model of other spared regions may counteract this bias. On the other hand, this study did not consider the problem of differential diagnosis with respect to other neurodegenerative diseases, which may give rise to cognitive impairment and progress to other forms of dementia such as fronto-temporal dementia (FTD), Parkinson disease or dementia with Lewy bodies. In order to support differential diagnosis, further studies must be developed to fit multi-class models to an extended FDG-PET dataset including patients suffering from different diseases. With this aim, the normalization criterion should be revised because cerebellar metabolism, by cross diaschisis, is more severely affected in other conditions such as FTD [44] and cerebellar normalization may deeply bias metabolic patterns. In the same condition, the inclusion of asymmetries and the combination of affected and unaffected regions may be even more important for the classification.

The present study, focused on an SVM-model based on FDG-PET data, was integrated with the explorative analysis of amyloid-deposition data. The analysis only involved the ADNI-set because amyloidosis biomarkers were available for few subjects in the GENOA-set. In fact the baseline work-up in this case series was collected several years ago (providing data with a long follow-up time), included FDG-PET, MRI and neuropsychology and was followed by repeated periodic clinical and neuropsychological examinations, but only the most recently enrolled patients underwent baseline amyloidosis biomarkers recording. Considering the ADNI-set, the rate of florbetapir positivity showed a pattern similar to the one reported in [35], around 30% positivity for NA, SMC and recovMCI patients, which can be viewed as limited specificity in the diagnostic process. On the other hand, the high rate of positivity (88%) in patients with AD diagnosis was close to the one found for the FDG-PET-based SVM model. The cross-analysis of FDG-PET based and amyloid-based classification showed that most of SVM-positive NA subjects were indeed florbetapir-positive, which could suggest a considerable risk of developing AD symptoms in the future. Analogously, about a half of SVM-negative subjects in the ADD group were also florbetapir-negative, which could

suggest the presence of some doubtful diagnoses. On the other hand, the majority of SVM-negative patients in the MCI groups were florbetapir-positive, which could be a limitation of the SVM-model but also suggests a possible delay in the transition from amyloidosis to neurodegeneration [45].

Conclusion

This study shows that a multivariate SVM model analyzing meta-VOI based FDG-PET data can reach high accuracy in classifying AD patients in both MCI and dementia stage and performs with generalization capability when applied to unknown independent data. Further training on suitable datasets can improve the sensitivity in the early stage of the disease. The automatic analysis of FDG-PET images may be integrated in a diagnostic process, which can take advantage of the analysis of multimodal data.

Acknowledgements Data collection and sharing for this project was partly funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, 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&Development, LLC.; Johnson & Johnson Pharmaceutical Research&Development LLC.; Lumosity; Lundbeck; Merck&Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research provides funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). 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. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

A substantial part of data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Compliance with ethical standards

Conflict of interest F.N.: received personal fees and nonfinancial support from GE Healthcare, non-financial support from Eli-Lilly and grants from Chiesi Farmaceutici; D.A. received speaking honoraria from Fidia s.p.a.; S.M. acted as a consultant for Eli Lilly in 2014 and for Avid Radiopharmaceuticals in 2016. All other authors have no potential conflicts to declare.

Ethical approval The institutional review board of the University of Genoa approved the recording and data treatment procedures involving human participants in this study and all procedures were in accordance with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Frisoni GB, Bocchetta M, Chételat G, Rabinovici GD, De Leon MJ, Kaye J, et al. Imaging markers for Alzheimer disease: which vs how. Neurology. 2013;81:487–500.
2. Jack CR, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol. 2010;9:119–28.
3. Landau SM, Harvey D, Madison CM, Koeppe RA, Reiman EM, Foster NL, et al. Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. Neurobiol Aging. 2011;32:1207–18.
4. Signorini M, Paulesu E, Friston K, Perani D, Colleluori A, Lucignani G, et al. Rapid assessment of regional cerebral metabolic abnormalities in single subjects with quantitative and nonquantitative [18F]FDG PET: a clinical validation of statistical parametric mapping. NeuroImage. 1999;9:63–80.
5. Caminiti SP, Ballarini T, Sala A, Cerami C, Presotto L, Santangelo R, et al. FDG-PET and CSF biomarker accuracy in prediction of conversion to different dementias in a large multicentre MCI cohort. NeuroImage Clin. Elsevier. 2018;18:167–77.
6. Perani D, Della Rosa PA, Cerami C, Gallivanone F, Fallanca F, Vanoli EG, et al. Validation of an optimized SPM procedure for FDG-PET in dementia diagnosis in a clinical setting. NeuroImage Clin [Internet]. Elsevier B.V. 2014;6:445–54. Available from: https://doi.org/10.1016/j.nicl.2014.10.009.
7. Herholz K, Salmon E, Perani D, Baron JC, Holthoff V, Frölich L, et al. Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET. NeuroImage. 2002;17:302–16.
8. Haense C, Herholz K, Jagust WJ, Heiss WD. Performance of FDG PET for detection of Alzheimer's disease in two independent multicentre samples (NEST-DD and ADNI). Dement Geriatr Cogn Disord. 2009;28:259–66.
9. Chen K, Ayutyanont N, Langbaum JBS, Fleisher AS, Reschke C, Lee W, et al. Characterizing Alzheimer's disease using a hypometabolic convergence index. NeuroImage [Internet]. Elsevier Inc. 2011;56:52–60. Available from: https://doi.org/10.1016/j.neuroimage.2011.01.049.
10. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage. 2002;15:273–89.
11. Pagani M, De Carli F, Morbelli S, Öberg J, Chincarini A, Frisoni GB, et al. Volume of interest-based [18F]fluorodeoxyglucose PET discriminates MCI converting to Alzheimer's disease from healthy controls. A European Alzheimer's disease consortium (EADC) study. NeuroImage Clin [Internet]. 2015;7:34–42. Available from: http://europepmc.org/abstract/med/25610765.
12. Pagani M, Nobili F, Morbelli S, Arnaldi D, Giuliani A, Öberg J, et al. Early identification of MCI converting to AD: a FDG PET study. Eur J Nucl Med Mol Imaging. 2017;44:2042–52.
13. Smailagic N, Vacante M, Hyde C, Martin S, Ukoumunne O, Sachpekidis C. F-FDG-PET for the early diagnosis of Alzheimer

' s disease dementia and other dementias in people with mild cognitive impairment (MCI) (review). Cochrane Database Syst Rev. 2015:1–104.

14. Rathore S, Habes M, Iftikhar MA, Shacklett A, Davatzikos C. A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. NeuroImage. Elsevier. 2017;155:530–48.
15. Morbelli S, Drzezga A, Perneczky R, Frisoni GB, Caroli A, Van Berckel BNM, et al. Resting metabolic connectivity in prodromal Alzheimer's disease. A European Alzheimer disease consortium (EADC) project. Neurobiol Aging [Internet]. Elsevier Inc. 2012;33:2533–50. Available from: https://doi.org/10.1016/j.neurobiolaging.2012.01.005.
16. Herholz K. PET studies in dementia. Ann Nucl Med. 2003;17:79–89.
17. Herholz K, Westwood S, Haense C, Dunn G. Evaluation of a calibrated 18F-FDG PET score as a biomarker for progression in Alzheimer disease and mild cognitive impairment. J Nucl Med [Internet]. 2011;52:1218–26. Available from: http://jnm.snmjournals.org/cgi/doi/10.2967/jnumed.111.090902.
18. Caroli A, Prestia A, Chen K, Ayutyanont N, Landau SM, Madison CM, et al. Summary metrics to assess Alzheimer disease-related hypometabolic pattern with 18F-FDG PET: head-to-head comparison. J Nucl Med [Internet]. 2012;53:592–600. Available from: http://jnm.snmjournals.org/cgi/doi/10.2967/jnumed.111.094946.
19. Pagani M, De Carli F, Morbelli S, Öberg J, Chincarini A, Frisoni GB, et al. Volume of interest-based [18F]fluorodeoxyglucose PET discriminates MCI converting to Alzheimer's disease from healthy controls. A European Alzheimer's disease consortium (EADC) study. NeuroImage Clin [Internet]. Elsevier B.V. 2015;7:34–42. Available from: https://doi.org/10.1016/j.nicl.2014.11.007.
20. Pagani M, Giuliani A, Öberg J, De Carli F, Morbelli S, Girtler N, et al. Progressive disintegration of brain networking from Normal aging to Alzheimer disease: analysis of independent components of 18 F-FDG PET data. J Nucl Med [Internet]. 2017;58:1132–9. Available from: http://jnm.snmjournals.org/lookup/doi/10.2967/jnumed.116.184309.
21. Petersen RC, Negash S. Mild cognitive impairment: an overview. CNS Spectr. 2008;13:45–53.
22. 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. Alzheimer's Dement [Internet]. Elsevier Ltd. 2011;7:270–9. Available from: https://doi.org/10.1016/j.jalz.2011.03.008.
23. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's Dement [Internet]. Elsevier Ltd. 2011;7:263–9. Available from: https://doi.org/10.1016/j.jalz.2011.03.005.
24. Varrone A, Asenbaum S, Vander Borght T, Booij J, Nobili F, Någren K, et al. EANM procedure guidelines for PET brain imaging using [18F]FDG, version 2. Eur J Nucl Med Mol Imaging. 2009;36:2103–10.
25. Della Rosa PA, Cerami C, Gallivanone F, Prestia A, Caroli A, Castiglioni I, et al. A standardized [18F]-FDG-PET template for spatial normalization in statistical parametric mapping of dementia. Neuroinformatics. 2014;12:575–93.
26. Soonawala D, Amin T, Ebmeier KP, Steele DJ, Dougall NJ, Best J, et al. Statistical parametric mapping of99mTc-HMPAO-SPECT images for the diagnosis of Alzheimer's disease: normalizing to cerebellar tracer uptake. NeuroImage. 2002;17:1193–202.
27. Dukart J, Mueller K, Horstmann A, Vogt B, Frisch S, Barthel H, et al. Differential effects of global and cerebellar normalization on detection and differentiation of dementia in FDG-PET studies. Neuroimage. Elsevier Inc. 2010;49:1490–5.
28. Morbelli S, Brugnolo A, Bossert I, Buschiazzo A, Frisoni GB, Galluzzi S, et al. Visual versus semi-quantitative analysis of18F-FDG-PET in amnestic MCI: an European Alzheimer's disease consortium (EADC) project. J Alzheimer's Dis [Internet]. 2015;44:815–26. Available from: http://europepmc.org/abstract/med/25362041.
29. Zahn R, Juengling F, Bubrowski P, Jost E, Dykierek P, Talazko J, et al. Hemispheric asymmetries of hypometabolism associated with semantic memory impairment in Alzheimer's disease: a study using positron emission tomography with fluorodeoxyglucose-F18. Psychiatry Res Neuroimaging. 2004;132:159–72.
30. Brown LD, Cai TT, Dasgupta A. Interval estimation for a binomial. Stat Sci. 2001;16:101–33.
31. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. J Clin Epidemiol [Internet]. 1991;44:763–70. Available from: http://www.sciencedirect.com/science/article/pii/089543569190128V.
32. Chen F, Xue Y, Tan MT, Chen P. Efficient statistical tests to compare Youden index: accounting for contingency correlation. Stat Med. 2015;34:1560–76.
33. Qin G, Hotilovac L. Comparison of non-parametric confidence intervals for the area under the ROC curve of a continuous-scale diagnostic test. Stat Methods Med Res. 2008;17:207–21.
34. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology [Internet]. Radiological Society of North America. 1982;143:29–36. Available from: https://doi.org/10.1148/radiology.143.1.7063747.
35. Landau SM, Mintun MA, Joshi AD, Koeppe RA, Petersen RC, Aisen PS, et al. Amyloid deposition, hypometabolism, and longitudinal cognitive decline. Ann Neurol [Internet]. 2012;72:578–86. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3786871/.
36. Cabral C, Morgado PM, Campos Costa D, Silveira M. Predicting conversion from MCI to AD with FDG-PET brain images at different prodromal stages. Comput Biol Med [Internet]. Elsevier. 2015;58:101–9. Available from: https://doi.org/10.1016/j.compbiomed.2015.01.003.
37. Pagani M, Giuliani A, Öberg J, Chincarini A, Morbelli S, Brugnolo A, et al. Predicting the transition from normal aging to Alzheimer's disease: a statistical mechanistic evaluation of FDG-PET data. Neuroimage [Internet]. Elsevier Inc. 2016;141:282–90. Available from: https://doi.org/10.1016/j.neuroimage.2016.07.043.
38. Arnaldi D, Morbelli S, Brugnolo A, Girtler N, Picco A, Ferrara M, et al. Functional neuroimaging and clinical features of drug naive patients with de novo Parkinson's disease and probable RBD. Parkinsonism Relat Disord. 2016;29:47–53.
39. Kryscio RJ, Abner EL, Cooper GE, Fardo DW, Jicha GA, Nelson PT, et al. Self-reported memory complaints: implications from a longitudinal cohort with autopsies. Neurology. 2014;83(15):1359-65.
40. Rasmussen JM, Lakatos A, van Erp TGM, Kruggel F, Keator DB, Fallon JT, et al. Empirical derivation of the reference region for computing diagnostic sensitive18fluorodeoxyglucose ratios in Alzheimer's disease based on the ADNI sample. Biochim Biophys Acta - Mol Basis Dis [Internet]. Elsevier B.V. 2012;1822:457–66. Available from: https://doi.org/10.1016/j.bbadis.2011.09.008.
41. Braak H, Braak E, Bohl J, Lang W. Alzheimer's disease: amyloid plaques in the cerebellum. J Neurol Sci. 1989;93:277–87.
42. Catafau AM, Bullich S, Seibyl JP, Barthel H, Ghetti B, Leverenz J, et al. Cerebellar amyloid- plaques: how frequent are they, and do they influence 18F-Florbetaben SUV ratios? J Nucl Med [Internet].

2016;57:1740–5 Available from: http://jnm.snmjournals.org/cgi/doi/10.2967/jnumed.115.171652.

43. Ishii K, Sasaki M, Kitagaki H, Yamaji S, Sakamoto S, Matsuda K, et al. Reduction of cerebellar glucose metabolism in advanced Alzheimer's disease. J Nucl Med [Internet]. 1997;38:925–8 Available from: http://jnm.snmjournals.org/content/38/6/925.full.pdf.
44. Bocchetta M, Cardoso MJ, Cash DM, Ourselin S, Warren JD, Rohrer JD. Patterns of regional cerebellar atrophy in genetic frontotemporal dementia. NeuroImage Clin [Internet]. The Authors. 2016;11:287–90. Available from: https://doi.org/10.1016/j.nicl.2016.02.008.
45. Bauckneht M, Chincarini A, Piva R, Arnaldi D, Girtler N, Massa F, et al. Metabolic correlates of reserve and resilience in MCI due to Alzheimer's disease (AD) Rik Ossenkoppele. Alzheimer's Res Ther Alzheimer's Research & Therapy. 2018;10:1–13.