SN Computer Science (2023) 4:249
https://doi.org/10.1007/s42979-023-01688-2

ORIGINAL RESEARCH

Integrating Different Data Modalities for the Classification of Alzheimer's Disease Stages

Lucia Maddalena[1] · Ilaria Granata[1] · Maurizio Giordano[1] · Mario Manzo[2] · Mario Rosario Guarracino[1,3]

Received: 25 May 2022 / Accepted: 11 January 2023 / Published online: 4 March 2023

Abstract

Alzheimer's disease (AD) is the most common form of dementia with physical, psychological, social, and economic impacts on patients, their carers, and society. Its early diagnosis allows clinicians to initiate the treatment as early as possible to arrest or slow down the disease progression more effectively. We consider the problem of classifying AD patients through a machine learning approach using different data modalities acquired by non-invasive techniques. We perform an extensive evaluation of a machine learning classification procedure using omics, imaging, and clinical features, extracted by the ANMerge dataset, taken alone or combined together. Experimental results suggest that integrating omics and imaging features leads to better performance than any of them taken separately. Moreover, clinical features consisting of just two cognitive test scores always lead to better performance than any of the other types of data or their combinations. Since these features are usually involved in the clinician diagnosis process, our results show how their adoption as classification features positively biases the results.

Keywords Data integration · Alzheimer's disease · Omics imaging · Transcriptomics · Magnetic resonance imaging

Introduction

Dementia is a syndrome that deteriorates cognitive functions beyond what might be expected from the usual consequences of biological aging. Currently, more than 55 million people live with dementia worldwide, and this number is envisioned to rise to 78 million in 2030, and 139 million in 2050 [1]. Alzheimer's disease (AD) is the most common form of dementia and may contribute to 60–70% of cases. It has physical, psychological, social, and economic impacts on people living with dementia, their carers, families, and society. Its early diagnosis allows clinicians to initiate the treatment as early as possible to arrest or slow down the disease progression more effectively [2].

Many machine learning approaches have been proposed for the automatic classification of AD stages, which go from cognitively normal (CN) to the intermediate state of mild cognitive impairment (MCI) to the final AD stage. Some exploit longitudinal studies to estimate the progression from one stage to the other [3–6], as reviewed in recent surveys [7–9]. Other approaches rely on cross-sectional studies, aiming to classify the degree of the disease based on the results of a predetermined visit [10–13]. Some of them [12–16] focus on traditional machine learning (ML) techniques and rely on hand-designed features extracted from

This article is part of the topical collection "Advances on Biomedical Engineering Systems and Technologies" guest edited by Hugo Gamboa, Ana Fred, Ana Roque, Denis Gracanin, Ronny Lorenz, Athanasios Tsanas and Nathalie Bier.

✉ Lucia Maddalena
lucia.maddalena@cnr.it

Ilaria Granata
ilaria.granata@cnr.it

Maurizio Giordano
maurizio.giordano@cnr.it

Mario Manzo
mmanzo@unior.it

Mario Rosario Guarracino
mariorosario.guarracino@unicas.it

[1] Institute for High-Performance Computing and Networking (ICAR), Consiglio Nazionale delle Ricerche (CNR), Via P. Castellino 111, 80131 Naples, Italy

[2] Information Technology Services, University of Naples "L'Orientale", Via Nuova Marina 59, 80133 Naples, Italy

[3] Department of Economics and Law, University of Cassino and Southern Lazio, Campus Folcara, 03043 Cassino, Italy

the data, according to domain-specific knowledge from AD research. Other methods [9, 11, 17–26] exploit the ability of deep learning (DL) architectures to discover the discriminant features in the data automatically. Examples of recent ML- and DL-based methods are given in "Comparison with the State-of-the-Art".

Biomedical image analysis has become a significant research field for various biomedical applications [27–33]. In the case of AD, the most frequently adopted imaging data include magnetic resonance images (MRIs) and positron emission tomography images (PETs) [9, 11, 17–19, 22, 34, 35]. Besides, other data modalities are commonly taken into account, including omics data (e.g., gene expression (GE) data [12, 36, 37]), sometimes coupled with clinical data [2, 6, 21, 38]. Recently, some research started focusing on the integration of omics data with information from biomedical images [13, 39–42]. These omics imaging methods, bringing together information from different sources, can reveal hidden genotype-phenotype relationships to understand the onset and progression of many diseases and identify new diagnostic and prognostic biomarkers [43].

In [13], we proposed an ML-based omics imaging approach to AD classification that relies on data acquired by non-invasive techniques. The multi-modal data, derived from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset,[1] consisted of omics and imaging features extracted by GE values from blood samples and MRIs, respectively. We have shown how a suitable integration of these data modalities, using well-known ML techniques, can often lead to better results than any of them taken separately.

Here, we explored a different dataset, namely the ANMerge dataset[2] [3], and performed a thorough performance analysis of AD classification results. Specifically, we varied the settings of the evaluation procedure based on cross-validation (CV), testing three different classifiers, four atlases for extracting MRI features, and different criteria for extracting the GE features. Analogous experiments have also been extended to data from ADNI to perform a fair comparison among the two sets of results. Finally, we also considered the adoption of clinical information (demographic features and cognitive tests), exploring the advantages of combining them with imaging and omics data types. As an added value, we make publicly available the software implementing the evaluation procedure for AD classification to simplify the comparison with results from other methods.

The rest of the paper is organized as follows. "Materials and Methods" describes the data adopted in the experiments for classifying AD patients and the procedures used to extract their features. "Experiments" presents the evaluation procedure and discusses the results achieved with the proposed framework. Finally, "Conclusions" concludes our paper and gives some future research directions.

Table 1 Number of ANMerge patients having MRI, GE, or both types of data, and their diagnosis (AD, MCI, or CN)

Data type	AD	MCI	CN	Total
MRI	128	131	116	375
GE	223	219	249	691
MRI+GE	102	112	103	317

Materials and Methods

The omics imaging data adopted in [13] for AD classification come from the ADNI database. For a detailed description of these data and the extraction of the related features, the reader can refer to [13]. Here, we considered analogous features obtained from the ANMerge dataset [3]. This dataset is an improved and updated version of AddNeuroMed [44]. It provides multi-modal data from more than 1,700 participants of a longitudinal study, including clinical assessments, MRIs, genotyping, transcriptomic profiling, and blood plasma proteomics.

We adopted GE values extracted by blood samples and MRIs, as we aim to use multi-modal information integrating omics and imaging data acquired by non-invasive techniques. In the ANMerge dataset, these data, available for selected subsets of patients, come both from the first visit, so they are already aligned in time. Table 1 summarizes the number of ANMerge patients having MRI, GE, or both types of data for their first visit and their diagnosis (AD, MCI, or CN). It can be observed that the three classes of the MRI+GE subset of ANMerge data selected for the experiments (in the following, simply referred to as the subset of ANMerge data) are very well balanced. This was not the case for the subset of ADNI data having both MRI and GE data used in [13], consisting of 42 AD, 428 MCI, and 250 CN patients.

Imaging Data: MRI

Rather than using the imaging features made available in ANMerge, to maintain consistency with [13], we extracted imaging features from MRIs using the Clinica open-source framework [9, 45, 46].

The basic step common to all Clinica pipelines for preprocessing and feature extraction involves the conversion of the dataset into the Brain Imaging Data Structure (BIDS) format [47]. Although automatic conversion tools are provided in Clinica for some publicly available datasets

[1] ADNI: https://adni.loni.usc.edu.

[2] ANMerge: https://doi.org/10.7303/syn22252881.

(including ADNI), there is no tool for ANMerge. Therefore, we prepared Matlab/Python scripts for centering the MRIs of all ANMerge selected patients and placing them into the BIDS structure.

Then, we adopted the Clinica framework to generate the voxel-based features from MRIs, which have been shown [9] to lead to high-performance results using the Support Vector Machine (SVM) [48] classifier. The Unified Segmentation procedure [49] is first applied, simultaneously performing tissue segmentation, bias correction, and spatial normalization of each input image. Next, a group template is created using the DARTEL algorithm for diffeomorphic image registration [50], using the subjects' tissue probability maps on the native space obtained from segmentation. The DARTEL to MNI method [50] is then applied, providing the registration of the native space images into the MNI (Montreal Neurological Institute) space, using routines from the Statistical Parametric Mapping[3] (SPM) package. These steps transform all the images into a common space, providing a voxel-wise correspondence across subjects. A set of imaging features is finally extracted based on regional measurements, where the anatomical regions are obtained from an atlas in the MNI space, and the average gray matter density is computed in each region. In the experiments, we considered the regional features derived from four different atlases: AAL2 [51] (121 features), AICHA [52] (385 features), Hammers [53, 54] (69 features), and LPBA40 [55] (57 features).

Omics Data: GE

Normalized gene expression data from blood transcriptomics of ANMerge participants were downloaded via Synapse.[4] Post-quality-control and batch-corrected expression values, as described in [3], were used for the differential abundance analysis. The lumiHumanIDMapping R package v. 1.10.1 [56] was used to map nuIDs to gene symbols. The dataset contained 5213 nuIDs for 691 under 90 samples. Multiple probes corresponding to the same gene symbol were aggregated by the median value. The Limma R package v. 3.46.0 [57] was used for performing differential expression analysis using linear models and finding significant differentially expressed genes (DEGs) from the three unpaired two-class contrast matrices (AD vs. CN, AD vs. MCI, MCI vs. CN). Several filtering criteria based on log-fold change (LogFC), which indicates the difference between two conditions in terms of gene expression, and the Benjamini–Hochberg adjusted p value (BH-adj.pvalue) were applied to each contrast to consider genes significant and select the omics features: (1) AD vs. CN: BH-adj. pvalue $\leq$ 0.05, 308 genes ("AD-CN_pv005"); (2) AD vs. MCI: BH-adj.pvalue $\leq$ 0.05, 6 genes ("AD-MCI_pv005") and BH-adj.pvalue $\leq$ 0.1, LogFC $\geq|$ 0.3 $|$, 59 genes ("AD-MCI_pv01_LFC03"); (3) MCI vs. CN: BH-adj.pvalue $\leq$ 0.05, LogFC $\geq|$ 0.3 $|$, 472 genes ("MCI-CN_pv005_LFC03") and BH-adj.pvalue $\leq$ 0.05, LogFC $\geq|$ 0.4 $|$, 42 genes ("MCI-CN_pv005_LFC04").

Table 2 Summary statistics describing the subset of ANMerge data

Features	AD (102)	MCI (112)	CN (103)	Total (317)
Sex				
Male # (%)	32 (31.37)	57 (50.89)	44 (42.72)	133 (41.96)
Female # (%)	70 (68.63)	55 (49.11)	59 (57.28)	184 (58.04)
Age (years)				
Mean (std)	75.45 (6.60)	73.96 (5.74)	72.54 (6.64)	73.98 (6.4)
Range	[58,88]	[56,86]	[52,87]	[52,88]
Education (years)				
Mean (std)	7.99 (3.98)	8.97 (4.29)	11.01 (4.88)	9.32 (4.56)
Range	[2,22]	[0,20]	[2,25]	[0,25]
MMSE				
Mean (std)	20.80 (4.67)	27.09 (1.72)	29.07 (1.2)	25.71 (4.54)
Range	[12,30]	[24,30]	[25,30]	[12,30]
CDR_SOB				
Mean (std)	6.66 (3.26)	1.43 (0.89)	0.07 (0.19)	2.67 (3.4)
Range	[0.5,15]	[0.5,4.5]	[0,1]	[0,15]

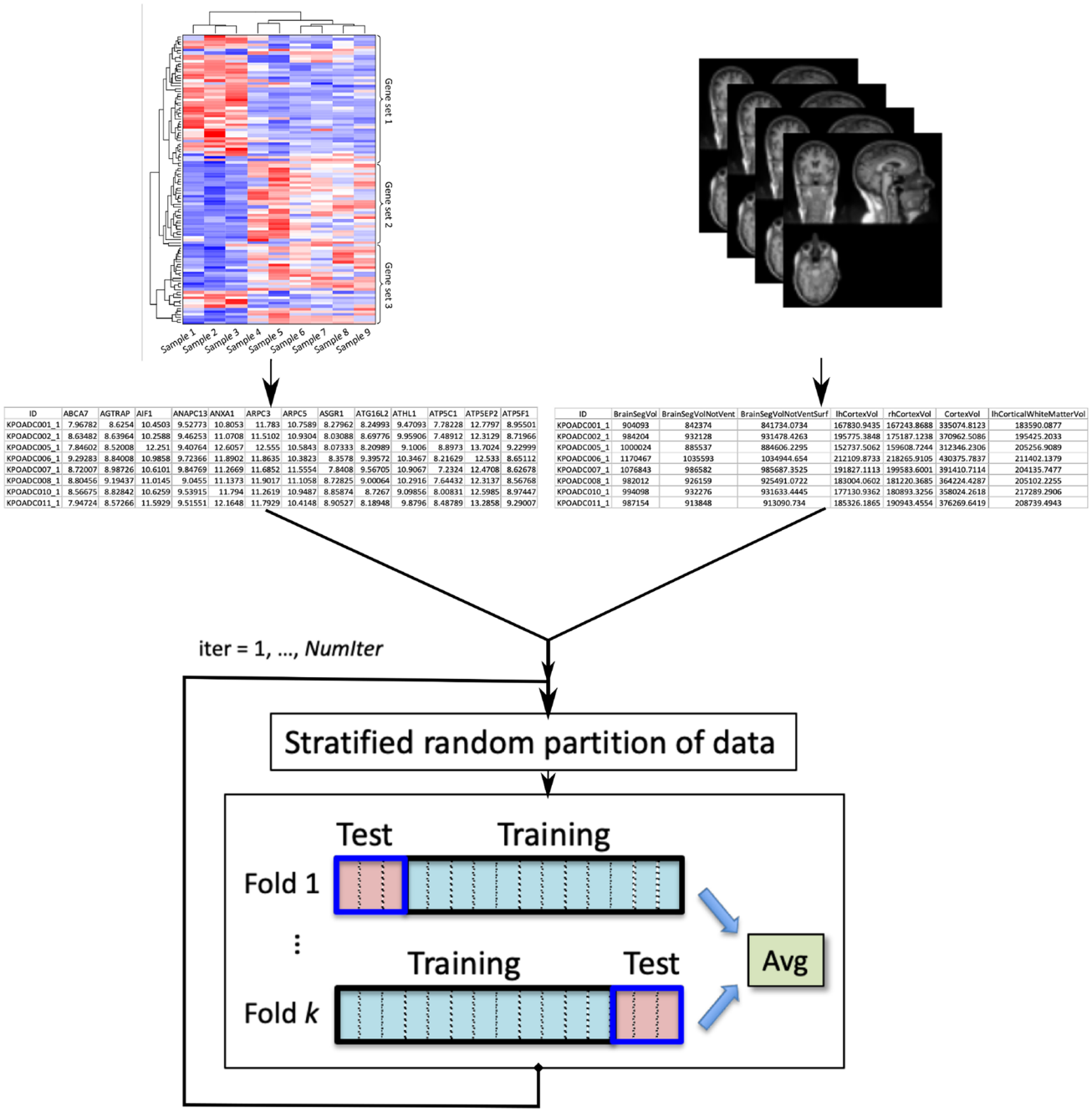
Clinical Data: Clin (Dem+Cog)

Often (e.g., [2, 6, 21, 38]) clinical data, eventually together with other types of data, are adopted as features for AD classification. As we already observed in [13], some of them are considered by medical doctors to diagnose the disease state of each patient, are directly adopted for labeling patients belonging to different classes [58], or are used as criteria for including/excluding patients from AD datasets, as in [3]. Therefore, their use as features for AD classification appears to bias the results strongly (and positively). Here, we investigated their role in classification performance, justifying our doubts.

Among the clinical information included in the ANMerge dataset, we selected five features that are available for most of the patients. These include three demographic features (sex, age, and years of education) and two cognitive test scores: the Clinical Dementia Rating Sum of Boxes (CDR_SOB) and the Mini-Mental State Examination (MMSE). Few missing values have been imputed

[3] SPM: https://www.fil.ion.ucl.ac.uk/spm/.

[4] Synapse: https://www.synapse.org.



ID	ABCA7	AGTRAP	AIF1	ANAPC13	ANXA1	ARPC3	ARPC5	ASGR1	ATG16L2	ATHL1	ATP5C1	ATP5EP2	ATP5F1
KPOADC001_1	7.96782	8.6254	10.4503	9.52773	10.8053	11.783	10.7589	8.27962	8.24993	9.47093	7.78228	12.7797	8.95501
KPOADC002_1	8.63482	8.63964	10.2588	9.46253	11.0708	11.5102	10.9304	8.03088	8.69776	9.95906	7.48912	12.3129	8.71966
KPOADC005_1	7.84602	8.52008	12.251	9.40764	12.6057	12.555	10.5843	8.07333	8.20989	9.1006	8.8973	13.7024	9.22999
KPOADC006_1	9.29283	8.84008	10.9858	9.72366	11.8902	11.8635	10.3823	8.3578	9.39572	10.3467	8.21629	12.533	8.65112
KPOADC007_1	8.72007	8.98726	10.6101	9.84769	11.2669	11.6852	11.5554	7.8408	9.56705	10.9067	7.2324	12.4708	8.62678
KPOADC008_1	8.80456	9.19437	11.0145	9.0455	11.1373	11.9017	11.1058	8.72825	9.00064	10.2916	7.64432	12.3137	8.56768
KPOADC010_1	8.56675	8.82842	10.6259	9.53915	11.794	11.2619	10.9487	8.85874	8.7267	9.09856	8.00831	12.5985	8.97447
KPOADC011_1	7.94724	8.57266	11.5929	9.51551	12.1648	11.7929	10.4148	8.90527	8.18948	9.8796	8.48789	13.2858	9.29007

ID	BrainSegVol	BrainSegVolNotVent	BrainSegVolNotVentSurf	lhCortexVol	rhCortexVol	CortexVol	lhCorticalWhiteMatterVol
KPOADC001_1	904093	842374	841734.0734	167830.9435	167243.8688	335074.8123	183590.0877
KPOADC002_1	984204	932128	931478.4263	195775.3848	175187.1238	370962.5086	195425.2033
KPOADC005_1	1000024	885537	884606.2295	152737.5062	159608.7244	312346.2306	205256.9089
KPOADC006_1	1170467	1035593	1034944.654	212109.8733	218265.9105	430375.7837	211402.1379
KPOADC007_1	1076843	986582	985687.3525	191827.1113	199583.6001	391410.7114	204135.7477
KPOADC008_1	982012	926159	925491.0722	183004.0602	181220.3685	364224.4287	205102.2255
KPOADC010_1	994098	932276	931633.4445	177130.9362	180893.3256	358024.2618	217289.2906
KPOADC011_1	987154	913848	913090.734	185326.1865	190943.4554	376269.6419	208739.4943

Fig. 1 Scheme of the evaluation procedure

through kNN separately for each class to introduce the minimum possible alteration of the data. The statistics of the subset of ANMerge data, extracted from the chosen clinical features, are reported in Table 2. In the experiments, we considered the results of using demographic features, cognitive tests, or both clinical data, also coupled with omics and imaging data.

Table 3 Best performance results on the subset of ANMerge data

	Acc	Sens	Spec	Prec	F1	Gm	AUC	MCC	BA
AD vs. CN									
MRI	0.813	0.828	0.798	0.808	0.815	0.811	0.879	0.631	0.813
GE	0.775	0.766	0.785	0.784	0.771	0.772	0.864	0.556	0.775
MRI+GE	0.874	0.872	0.875	0.879	0.872	0.872	0.949	**0.752**	**0.873**
AD vs. MCI									
MRI	0.707	0.707	0.707	0.693	0.695	0.702	0.771	0.419	0.707
GE	0.588	0.539	0.633	0.577	0.553	0.579	0.636	0.176	0.586
MRI+GE	0.722	0.720	0.724	0.710	0.710	0.717	0.791	**0.449**	**0.722**
MCI vs CN									
MRI	0.611	0.613	0.609	0.634	0.619	0.606	0.650	0.225	0.611
GE	0.746	0.766	0.725	0.757	0.758	0.741	0.828	0.496	0.745
MRI+GE	0.778	0.790	0.765	0.790	0.786	0.774	0.855	**0.560**	**0.777**

In bold the best MCC and BA results for each binary classification task

Experiments

Evaluation Procedure

The evaluation procedure adopted in the experiments, illustrated in Fig. 1, is similar to the one adopted in [13].

It consists of *NumIter* iterations of *k*-fold cross-validation, with stratified partitions of the data into training and test subsets, using different classifiers. At each iteration, data are standardized by z-scoring training folds and using their mean and variance to z-score the test fold accordingly. The performance results are computed as average over the *NumIter* iterations of a set of well-known metrics: accuracy (Acc), sensitivity (Sens), specificity (Spec), precision (Prec), F-measure (F1), geometric mean (Gm), area under the ROC curve (AUC), Matthews correlation coefficient (MCC), and balanced accuracy (BA). For their description, please, refer to [13]. The reason for using so many metrics is to make the comparison of other methods with our results easier. However, most of our conclusions are based on values achieved for MCC and BA, which provide clear overall results regardless of eventual class unbalancing [59, 60]. The Matlab scripts implementing the evaluation procedure and the computation of the performance metrics are made publicly available through our web pages.

Evaluation on ANMerge Data

The described procedure has been applied to the subset of ANMerge data for each of the three binary problems. We fixed to 5 the number of folds for the CV and to 50 the number of iterations of the CV and varied the classifier (SVM with linear kernel, kNN with k=5, and logistic regression (LR)), the atlas for imaging features (AAL2, AICHA, Hammers, and LPBA40, see "Imaging Data: MRI"), and the filtering criterion for extracting the GE features (AD-CN_pv005 for AD vs. CN, AD-MCI_pv005 and AD-MCI_pv01_LFC03 for AD vs. MCI, MCI-CN_pv005_LFC03 and MCI-CN_pv005_LFC04 for MCI vs. CN, see "Omics Data: GE").

The best performance results for each of the ANMerge binary problems are reported in Table 3. They are those showing the highest MCC and BA values, achieved using almost always the SVM classifier except for all the cases of MRI data alone and the MRI+GE data for the AD vs. MCI task, where LR leads to better performance. The atlas for MRI data was AAL2 in all cases, except for the MRI+GE data for the AD vs. MCI task, where LPBA40 leads to better performance. The chosen criteria for GE data were the AD-CN_pv005 for the AD vs. CN task, the AD-MCI_pv005 for the AD vs. MCI task, and the MCI-CN_pv005_LFC03 for the MCI vs. CN task.

Based on the described best choices for setting up the evaluation procedure, Table 3 suggests that using omics imaging data (MRI+GE) from the subset of ANMerge data leads to better classification results than imaging or omics data alone for all three binary classification tasks.

Comparison with Results on ADNI Data

To better compare the results achieved on the subset of ANMerge data with those that can be obtained using the ADNI dataset, we repeated the experiments reported in [13], but fixing the same parameter values as above (5 folds and 50 iterations) and varying the classifier and the atlas as done for the ANMerge dataset. The method for extracting the GE features varied among those described in [13] (SAM-Tstat and SAM-Wilc for both AD vs. CN and AD vs. MCI and Topvar300 for MCI vs. CN). The best performance results are reported in Table 4 for each of the ADNI binary problems. The highest MCC and BA values have been reached using the SVM classifier, except for the GE and MRI+GE

Table 4 Best performance results on the subset of ADNI data

	Acc	Sens	Spec	Prec	F1	Gm	AUC	MCC	BA
AD vs. CN									
MRI	0.922	0.605	0.975	0.815	0.681	0.760	0.916	0.655	0.790
GE	0.853	0.476	0.916	0.508	0.476	0.649	0.835	0.401	0.696
MRI+GE	0.941	0.688	0.984	0.883	0.764	0.817	0.951	**0.745**	**0.836**
AD vs. MCI									
MRI	0.884	0.242	0.947	0.320	0.264	0.448	0.773	0.212	0.595
GE	0.887	0.359	0.939	0.370	0.353	0.560	0.771	0.298	0.649
MRI+GE	0.918	0.454	0.963	0.560	0.487	0.649	0.880	**0.454**	**0.709**
MCI vs. CN									
MRI	0.632	0.732	0.462	0.700	0.715	0.579	0.645	**0.198**	**0.597**
GE	0.591	0.851	0.145	0.630	0.719	0.253	0.514	−0.002	0.498
MRI+GE	0.617	0.753	0.384	0.677	0.712	0.530	0.614	0.145	0.568

In bold the best MCC and BA results for each binary classification task

cases in the MCI vs. CN task, where LR leads to better results. As in [13], the best choices for the AD vs. CN task are confirmed to be the AICHA atlas and the SAM-Wilc method, and the same can be said for the MCI vs. CN task with the AAL2 atlas and the Topvar300 method. However, for the AD vs. MCI task, the best atlas was LPBA40 (instead of AICHA), and the best GE extraction method was SAM-Tstat (instead of SAM-Wilc).

The overall results obtained in [13] for the ADNI data were confirmed, as omics imaging data allow us to achieve better performance than imaging or omics data alone for all the binary tasks, except the MCI vs. CN. Besides the latter task being much more challenging than the others, worse results on this dataset (but not on the ANMerge dataset) are probably due to the more unbalanced class distribution.

Apart from different choices for the MRI atlas and the GE extraction procedure, performance results on ADNI data in terms of MCC and BA appear to be similar to those on ANMerge data for the AD vs. CN and AD vs. MCI tasks. However, probably due to the strong class unbalancing for these tasks in the subset of ADNI data, the accuracy for the AD minority class (i.e., Sens) is much lower than that of the other two (majority) classes (i.e., Spec). This phenomenon is absent in the results on the well-balanced subset of ANMerge data, where Sens and Spec achieve similar results.

Instead, a significant performance improvement can be observed when using data from ANMerge for the MCI vs. CN task, which was an open issue already highlighted in [13]. This appears to be connected to more significant GE features that succeeded in being extracted compared to those from the ADNI data.

Evaluation Using Also Clinical Data

Best classification results obtained using demographic features (Dem), cognitive tests (Cog), or all Clinical (Dem+Cog) data, also in combination with MRI and GE data, are reported in Table 5. The best performance values have been chosen by fixing the SVM classifier while varying the MRI atlas and the GE criterion.

From Table 5, we can observe that clinical data alone, mainly just the two cognitive test scores, always lead to better performance than any of the other types of data or their combinations. This is undoubtedly due to their strong link with the medical doctor's diagnosis and justifies our concern in using these features for a fair evaluation. Indeed, cognitive test scores confirm being strongly informative of patients' AD status. However, we believe these tests alone cannot solve the problem, as patients get acquainted with them over time, and their scores, in subsequent visits, would be biased by the acquired experience. Therefore, an automatic tool for AD classification based on omics imaging data acquired by non-invasive techniques, such as the one proposed, could help clinicians make their diagnoses.

Comparison with the State-of-the-Art

Being ANMerge relatively recent, we could not find other methods using the dataset to evaluate AD classification in cross-sectional studies. Nonetheless, to provide a rough performance comparison of the results that could be achieved by state-of-the-art approaches, in Table 6, we report the classification performance of several methods on various datasets published in the recent literature. Our best results from Table 3 are also reported to make more immediate comparisons. For each classification problem, methods are grouped as DL-based (top) and ML-based (bottom). For each method, we report an acronym (made by the last name of the first author and the publication year) and reference, the adopted dataset, the number of samples for each class, the type(s) of data

Table 5 Best performance results on the subset of ANMerge data using also clinical data

	Acc	Sens	Spec	Prec	F1	Gm	AUC	MCC	BA
AD vs. CN									
Dem	0.635	0.698	0.573	0.622	0.654	0.628	0.690	0.277	0.636
Cog	0.985	0.969	1.000	1.000	0.984	0.984	0.998	**0.970**	**0.984**
Clinical	0.984	0.968	1.000	1.000	0.984	0.984	0.998	0.969	0.984
MRI+Dem	0.787	0.779	0.795	0.797	0.784	0.784	0.853	0.579	0.787
MRI+Cog	0.960	0.945	0.974	0.974	0.959	0.959	0.989	0.921	0.959
MRI+Clin	0.959	0.946	0.971	0.972	0.958	0.958	0.988	0.920	0.959
GE+Dem	0.804	0.806	0.803	0.807	0.803	0.802	0.895	0.614	0.804
GE+Cog	0.916	0.895	0.937	0.937	0.913	0.915	0.976	0.835	0.916
GE+Clin	0.922	0.904	0.941	0.941	0.920	0.921	0.980	0.848	0.922
MRI+GE+Dem	0.873	0.860	0.885	0.886	0.870	0.871	0.948	0.750	0.873
MRI+GE+Cog	0.929	0.923	0.934	0.935	0.928	0.928	0.984	0.860	0.929
MRI+GE+Clin	0.928	0.922	0.933	0.934	0.927	0.927	0.983	0.858	0.928
AD vs. MCI									
Dem	0.586	0.668	0.512	0.557	0.604	0.579	0.606	0.184	0.590
Cog	0.932	0.902	0.959	0.955	0.926	0.930	0.967	**0.867**	**0.931**
Clinical	0.921	0.879	0.959	0.953	0.913	0.917	0.968	0.845	0.919
MRI+Dem	0.676	0.656	0.695	0.667	0.657	0.671	0.733	0.356	0.676
MRI+Cog	0.886	0.885	0.887	0.882	0.881	0.885	0.946	0.776	0.886
MRI+Clin	0.882	0.880	0.884	0.878	0.877	0.881	0.947	0.768	0.882
GE+Dem	0.655	0.609	0.697	0.651	0.625	0.647	0.717	0.311	0.653
GE+Cog	0.927	0.891	0.959	0.954	0.920	0.924	0.968	0.857	0.925
GE+Clin	0.912	0.886	0.935	0.929	0.905	0.909	0.965	0.827	0.911
MRI+GE+Dem	0.710	0.685	0.734	0.708	0.691	0.705	0.782	0.424	0.709
MRI+GE+Cog	0.902	0.891	0.913	0.907	0.896	0.900	0.961	0.808	0.902
MRI+GE+Clin	0.893	0.877	0.908	0.900	0.886	0.891	0.959	0.790	0.893
MCI vs CN									
Dem	0.587	0.697	0.468	0.588	0.636	0.567	0.615	0.170	0.582
Cog	0.936	0.974	0.895	0.915	0.942	0.932	0.988	0.879	0.935
Clinical	0.947	0.973	0.918	0.932	0.951	0.944	0.989	**0.898**	**0.946**
MRI+Dem	0.581	0.597	0.565	0.602	0.595	0.574	0.602	0.164	0.581
MRI+Cog	0.880	0.876	0.885	0.896	0.883	0.879	0.953	0.765	0.881
MRI+Clin	0.884	0.890	0.877	0.891	0.888	0.882	0.956	0.772	0.883
GE+Dem	0.750	0.769	0.729	0.759	0.761	0.745	0.830	0.503	0.749
GE+Cog	0.916	0.918	0.915	0.924	0.919	0.915	0.978	0.835	0.916
GE+Clin	0.916	0.917	0.915	0.924	0.919	0.915	0.976	0.835	0.916
MRI+GE+Dem	0.781	0.795	0.766	0.791	0.790	0.778	0.860	0.566	0.780
MRI+GE+Cog	0.896	0.916	0.875	0.892	0.902	0.894	0.956	0.796	0.896
MRI+GE+Clin	0.893	0.913	0.872	0.889	0.899	0.891	0.952	0.790	0.892

In bold the best MCC and BA results for each binary classification task

modality, and the performance values in terms of the most commonly used metrics (Acc, Sens, AUC, and BA).

DL-based methods considered in Table 6 include [9, 11, 17–26]. Aderghal et al. [17] integrate the MRI and DTI (Diffusion Tensor Imaging) modalities from ADNI data. Due to the scarcity of DTIs, they adopt cross-modal transfer learning from MRIs to DTIs and combine the classification results of multiple CNNs by a majority vote. In [18], Backstrom et al. propose a 3D CNN for AD vs. CN classification using ADNI MRIs. Multiple MRIs for individual subjects, made at different times, are considered; the results reported in Table 6 are those obtained by the authors using a subject-separated data partitioning strategy. Li et al. [19] propose a classification method based on multiple cluster dense convolutional neural networks (DenseNets) to learn features from MRIs coming from ADNI. Each whole-brain image is first partitioned into different local regions, and a fixed number of 3D patches is extracted from each region. These

Table 6 Performance (%) on various datasets of recent methods for AD classification

Acronym [Ref.]	Year	Dataset	# Samples	Feats.	Acc	Sens	Spec	Prec	F1	AUC	BA
AD vs. CN											
Aderghal2018 [17]	2018	ADNI	48, 58	MRI, DTI	92.5	94.7	90.4	–	–	–	92.5
Backstrom2018 [18]	2018	ADNI	199, 141	MRI	90.1	93.3	86.8	–	–	–	90.0
Li2018 [19]	2018	ADNI	199, 229	MRI	89.5	87.9	90.8	–	–	92.4	89.4
Pan2018 [20]	2018	ADNI	428, 205	MRI, PET	92.5	89.9	94.5	–	91.3	85.9	92.2
Senanayake2018 [21]	2018	ADNI	161, 161	MRI, Cog	78.0	–	–	–	–	–	–
Shi2018 [22]	2018	ADNI	51, 52	MRI, PET	97.1	95.9	98.5	–	–	–	97.2
Lazni2019 [61]	2019	ADNI	77,82	MRI, PET	93.6	91.4	–	–			
Stamate2019 [23]	2019	EMIF-AD	115, 242	Blood metab		–	–	–	–	88.0	–
Bae2020 [11]	2020	ADNI	195, 195	MRI	89.0	88.0	91.0	–	–	94.0	89.5
Bae2020 [11]	2020	SNUBH	195, 195	MRI	88.0	85.0	91.0	–	–	91.0	88.0
Islam2020 [24]	2020	ADNI	98, 105	PET	71.5	–	–	–	–	–	–
Jo2020 [25]	2020	ADNI	66, 66	PET	90.8	–	–	–	–	–	–
Wen2020 [9]	2020	ADNI	236, 230	MRI	–	–	–	–	–	–	89.0
Wen2020 [9]	2020	AIBL	76, 429	MRI	–	–	–	–	–	–	88.0
Wen2020 [9]	2020	OASIS	78, 76	MRI	–	–	–	–	–	–	70.0
Yu2022 [26]	2020	ADNI	221, 315	MRI	95.9	95.8	–	–	–	95.9	–
Hett2018 [14]	2018	ADNI	186, 226	MRI	91.3	93.4	87.6	–	–	94.7	90.5
Zheng2018 [15]	2018	ADNI	142, 165	MRI	98.7	98.6	98.8	–	–	**99.9**	**98.7**
Gupta2019 [16]	2019	ADNI	38, 38	MRI, PET, APOE, CSF	98.4	100	96.5	97.9	98.4	<u>98.3</u>	<u>98.3</u>
Lee2020 [12]	2020	ADNI	63, 136	GE	–	–	–	–	–	65.7	–
Lee2020 [12]	2020	ANM1	145, 104	GE	–	–	–	–	–	87.4	–
Lee2020 [12]	2020	ANM2	139, 134	GE	–	–	–	–	–	80.4	–
Maddalena2022 [13]	2022	ADNI	42, 250	MRI, GE	94.6	72.2	98.3	88.9	78.7	95.5	85.3
Ours	2022	ANMerge	102, 103	MRI, GE	87.4	87.2	87.5	87.9	87.2	94.9	87.3
AD vs. MCI											
Aderghal2018 [17]	2018	ADNI	48, 108	MRI, DTI	85.0	93.7	79.1	–	–	–	**86.4**
Senanayake2018 [21]	2018	ADNI	161, 193	MRI, Cog	76.0	–	–	–	–	–	–
Yu2022 [26]	2022	ADNI	221, 297	MRI	85.7	88.9	–	–	–	<u>85.4</u>	–
Zheng2018 [15]	2018	ADNI	142, 221	MRI	73.8	64.1	80.1	–	–	77.3	72.1
Maddalena2022 [13]	2022	ADNI	42, 428	MRI, GE	91.5	39.4	96.6	54.6	44.8	**86.9**	68.0
Ours	2022	ANMerge	102, 112	MRI, GE	72.2	72.0	72.4	71.0	71.0	79.1	<u>72.2</u>
MCI vs. CN											
Aderghal2018 [17]	2018	ADNI	108, 58	MRI, DTI	80.0	92.8	73.0	–	–	–	<u>82.9</u>
Li2018 [19]	2018	ADNI	403, 229	MRI	73.8	86.6	51.5	–	–	77.5	80.2
Senanayake2018 [21]	2018	ADNI	193, 161	MRI, Cog	75.0	–	–	–	–	–	–
Shi2018 [22]	2018	ADNI	99, 52	MRI, PET	87.2	97.9	67.0	–	–	–	82.5
Yu2022 [26]	2022	ADNI	297, 315	MRI	89.3	96.7	–	–	–	<u>88.7</u>	–
Zheng2018 [15]	2018	ADNI	221, 165	MRI	97.9	98.6	97.0	–	–	**99.3**	**97.8**
Maddalena2022 [13]	2022	ADNI	428, 250	MRI	63.6	73.2	47.1	70.4	71.7	65.1	60.1
Ours	2022	ANMerge	112, 103	MRI, GE	77.8	79.0	76.5	79.0	78.6	85.5	77.7

For each classification problem, methods are grouped as DL- (top) and traditional ML-based (bottom). For each method, we report our acronym and reference (Acronym [Ref.]), the adopted dataset (Dataset), the number of samples for each class (# Samples), the type(s) of data modality (Feats.), and the performance values. In boldface/underlined the best/second best AUC and BA values for each classification problem

patches are grouped into different clusters with k-means clustering, and a DenseNet is constructed to learn the patch features for each cluster. The features learned from the discriminating clusters of each region are combined for classification, and the results from different local regions are integrated to enhance the final image classification. Pan

et al. [20] propose a DL framework for AD diagnosis based on MRI and PET images from the ADNI dataset. Missing PET images are imputed from their corresponding MRI data by using 3D Cycle-consistent Generative Adversarial Networks (3D-cGAN) to capture their relationship. A deep multi-modal multi-instance neural network is then used for AD classification using subjects with both MRI and PET (either real or synthetic). Senanayake et al. [21] use 3D MR volumes and neuropsychological measure-based (NM) feature vectors from ADNI. To combine these two data sources, having very different dimensions (35 NM features against more than ten million features from 3D MR volumes), they propose a DL-based pipeline that reduces the dimension of the MRI features to a dimension comparable with that of NM, and use the feature vector merging the two sets of features. The accuracy values reported in Table 6 have been extracted by their bar plots. Shi et al. [22] propose a multimodal algorithm based on a stacked deep polynomial network (MM-SDPN). Two SDPNs are first used to learn high-level features from MRIs and PETs coming from the ADNI dataset, taken separately. These are then fed to another SDPN to fuse multi-modal neuroimaging information to retain the intrinsic properties of both modalities and their correlation. Bae et al. [11] develop a CNN-based algorithm to classify AD patients using coronal slices of T1-weighted MRIs that cover the medial temporal lobe. The performance results reported in Table 6 come from the within-dataset validation performed by the authors on data from the ADNI and the Seoul National University Bundang Hospital (SNUBH) datasets. Islam et al. [24] propose an approach for generating synthetic brain PET images for building a large-scale dataset for training DL models for AD classification. The PET images generator exploits Deep Convolutional Generative Adversarial Networks (DCGANs) [62] learning on the three AD stages. A 2D-CNN model using axial, coronal, and sagittal slices from the generated PET data is finally adopted for classification. Jo et al. [25] propose a 3D CNN-based DL model on PET images for AD classification. Random under-sampling (RUS) is adopted for class balancing. Wen et al. [9] present an open-source framework for AD classification using CNNs and T1-weighted MRIs and use it to compare different CNN architectures. Table 6 reports the best results obtained by training the DL models on a subset of ADNI data and testing them on the remaining ADNI data, as well as on the AIBL[5] (the Australian Imaging, Biomarkers and Lifestyle Flagship Study of Ageing) and OASIS[6] (Open Access Series of Imaging Studies) datasets. Yu et al. [26] propose the THS-GAN (Tensor-train decomposition, Higher-order pooling, and Semi-supervised learning-based GAN) method for AD classification on ADNI data. The tensor-train decomposition is applied to all layers in the classifier and discriminator, reducing the number of parameters and exploiting, at the same time, the structural information of the brain. The higher-order pooling leverages the second-order statistics of the MRIs, effectively capturing long-range dependencies between slices of different directions. Moreover, the model is designed in a semi-supervised manner to take advantage of both labeled and unlabeled MRIs.

ML-based methods considered in Table 6 include [12–16]. Hett et al. [14] propose a texture-based grading framework based on 3D Gabor filters to better capture structural alterations caused by AD. An adaptive patch-based fusion strategy based on a local confidence criterion is adopted to combine all the grading maps estimated on texture maps. Moreover, contrary to usual grading-based methods using the average grading values over the considered ROI, a classification step is applied based on a non-parametric grading values distribution representation to better discriminate the pathology stages. Zheng et al. [15] describe a network-based approach built on multiple morphological features to enhance the MRI-based accuracy of AD classification. A multifeature-based network (MFN) is constructed for each patient using a sparse linear regression performed on six types of morphological features to promote the structure-based diagnosis. SVM was adopted to examine the diagnostic performance of the MFN by cross-validating the results. The performance values reported in Table 6 were obtained by combining the properties of the MFN with the morphological features. Gupta et al. [16] propose a framework based on SVM and feature selection to discriminate the various stages of ADNI patients using a combination of FDG-PETs, MRIs, CSF protein levels, and APOE genotype data. Stamate et al. [23] evaluate the performance of three state-of-the-art machine learning models (XGBoost, RandomForest, and DL) on AD classification based on plasma metabolites as potential AD biomarkers. Data samples are gathered by the European Medical Information Framework for AD Multimodal Biomarker Discovery[7] (EMIF-AD) [63]. The study demonstrates that XGBoost, whose performance is reported in Table 6, is more effective than RF and DL for this particular dataset and that this accuracy for clinical diagnosis is broadly similar to that achieved by CSF markers of AD pathology [64]. Lee and Lee [12] classify AD vs. CN using blood gene expression data not only from the ADNI dataset but also from two AddNeuroMed[8] [44] datasets (ANM1 and ANM2). Table 6 reports the best results obtained using suitable feature selection methods and classifiers.

[5] AIBL: https://aibl.csiro.au.

[6] OASIS: https://www.oasis-brains.org.

[7] EMIF-AD: http://www.emif.eu.

[8] AddNeuroMed: https://www.synapse.org/#!Synapse:syn4907804.

Despite the variability of experimental conditions and, consequently, of the performance results, few observations can be derived from Table 6. (1) Generally, but not always (e.g., see the results of [65] and [15]), DL-based methods lead to better performance than those based on traditional ML; (2) Different datasets can lead to pretty different performances of the same method (e.g., see the results of [9] and [12] varying the dataset); (3) The combination of multiple data sources appears promising for better performance (e.g., [17, 20, 22]); however, it should be taken into account that the extraction of further data modalities (e.g., the CSF used in [16]) could require quite invasive interventions; (4) Our performance results using MRI and GE features appear on average compared to those achieved by state-of-the-art methods for various classification tasks. Being the only ones produced for the ANMerge dataset, they can be used as a baseline for evaluating future AD classification methods on different data. Moreover, the promising integration of different data modalities obtainable by non-invasive techniques could be exploited by future ML- or DL-based methods.

Conclusions

We proposed an extensive evaluation of a machine learning procedure for classifying Alzheimer's patients using data from the ANMerge dataset. We considered data from different modalities, including imaging, omics, and clinical features, taken alone or combined together.

Overall results suggest that integrating omics and imaging features leads to better performance than any of them taken separately. This result holds whichever is the binary AD classification problem taken into account, also for the most challenging task of distinguishing MCI vs. CN patients, differently from what we previously experimented on analogous data from the ADNI dataset.

Moreover, we showed that clinical features consisting of just two cognitive test scores always lead to better performance than any other types of data or their combinations. Our results show how their adoption as classification features positively biases the results, being involved in the clinician diagnosis process, thus inhibiting a fair evaluation.

We believe that the results of our extensive experiments on the ANMerge dataset can be used as a baseline to compare for the evaluation of future AD classification methods. Indeed, as the dataset (at least in its merged and revised form) is pretty new, no other methods could be found experiencing it. Toward this goal, we believe that the public availability of the software developed for the experiments makes these future comparisons easier.

Future research is directed toward adopting different cohort study data for independent training and testing. This will certainly require a general method for reducing batch effects between different experiments and an *ad hoc* procedure for classifier hyperparameter optimization on the training set.

Supplementary information

The Matlab scripts implementing the evaluation procedure adopted for AD classification and the computation of the performance metrics are made publicly available through our web pages.

Acknowledgements This work has been partially funded by the BiBiNet project (H35F21000430002) within POR-Lazio FESR 2014–2020. It was carried out also within the activities of the authors as members of the INdAM Research group GNCS and the ICAR-CNR INdAM Research Unit and partially supported by the INdAM research project "Computational Intelligence methods for Digital Health". The work of Mario R. Guarracino was conducted within the framework of the Basic Research Program at the National Research University Higher School of Economics (HSE). Mario Manzo thanks Prof. Alfredo Petrosino for his guidance and supervision during the years of working together.

Funding Open access funding provided by Consiglio Nazionale Delle Ricerche (CNR) within the CRUI-CARE Agreement.

Data availability The results published here are based on ANMerge data obtained from the AD Knowledge Portal (https://acknowledgeportal.org) and on data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (https://adni.loni.usc.edu).

Declarations

Conflict of interest The authors declare that they have no known conflict of interests.

References

1. World Health Organization: Dementia Key Factors (2021). https://www.who.int/en/news-room/fact-sheets/detail/dementia
2. Bucholc M, Ding X, Wang H, Glass DH, Wang H, Prasad G, Maguire LP, Bjourson AJ, McClean PL, Todd S, Finn DP, Wong-Lin K. A practical computerized decision support system for predicting the severity of Alzheimer's disease of an individual. Expert Syst Appl. 2019;130:157–71. https://doi.org/10.1016/j.eswa.2019.04.022.
3. Birkenbihl C, Westwood S, Shi L, Nevado-Holgado A, Westman E, Lovestone S, Consortium A, Hofmann-Apitius M. ANMerge: a comprehensive and accessible Alzheimer's disease patient-level dataset. J Alzheimers Dis. 2021;79:423–31. https://doi.org/10.3233/JAD-200948.

4. Lovestone S, Francis P, Kloszewska I, Mecocci P, Simmons A, Soininen H, Spenger C, Tsolaki M, Vellas B, Wahlund L, Ward M, Consortium A. AddNeuroMed—the European collaboration for the discovery of novel biomarkers for Alzheimer's disease. Ann N Y Acad Sci. 2009. https://doi.org/10.1111/j.1749-6632.2009.05064.x.
5. Mueller S, Weiner M, Thal L, Petersen R, Jack C, Jagust W, Trojanowski J, Toga A, Beckett L. Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). J Alzheimers Dement. 2005;1(1):55–66. https://doi.org/10.1016/j.jalz.2005.06.003.
6. Spasov SE, Passamonti L, Duggento A, Liò P, Toschi N. A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease. Neuroimage. 2019;189:276–87. https://doi.org/10.1016/j.neuroimage.2019.01.031.
7. Lawrence E, Vegvari C, Ower A, Hadjichrysanthou C, De Wolf F, RM A. A systematic review of longitudinal studies which measure Alzheimer's disease biomarkers. J Alzheimers Dis. 2017;59(4):1359–79. https://doi.org/10.3233/JAD-170261.
8. Martí-Juan G, Sanroma-Guell G, Piella G. A survey on machine and statistical learning for longitudinal analysis of neuroimaging data in Alzheimer's disease. Comput Methods Programs Biomed. 2020;189: 105348. https://doi.org/10.1016/j.cmpb.2020.105348.
9. Wen J, Thibeau-Sutre E, Diaz-Melo M, Samper-González J, Routier A, Bottani S, Dormont D, Durrleman S, Burgos N, Colliot O. Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation. Med Image Anal. 2020;63: 101694. https://doi.org/10.1016/j.media.2020.101694.
10. Fiscon G, Weitschek E, Cialini A, et al. Combining EEG signal processing with supervised methods for Alzheimer's patients classification. BMC Med Inform Decis Mak. 2018. https://doi.org/10.1186/s12911-018-0613-y.
11. Bae J, Lee S, Jung W, Park S, Kim W, Oh H, Han J, Kim G, Kim J, Kim J, Kim K. Identification of Alzheimer's disease using a convolutional neural network model based on T1-weighted magnetic resonance imaging. Sci Rep. 2020. https://doi.org/10.1038/s41598-020-79243-9.
12. Lee T, Lee H. Prediction of Alzheimer's disease using blood gene expression data. Sci Rep. 2020;10(1):3485. https://doi.org/10.1038/s41598-020-60595-1.
13. Maddalena L, Granata I, Giordano M, Manzo M, Guarracino MR. Alzheimer's disease neuroimaging initiative (ADNI): classifying Alzheimer's disease using mris and transcriptomic data. In: Proceedings of the 15th international joint conference on biomedical engineering systems and technologies—BIOIMAGING. Portugal: SciTePress. 2022. p. 70–9. https://doi.org/10.5220/0010902900003123. INSTICC.
14. Hett K, Ta V-T, Manjón JV, Coupé P, Initiative ADN, et al. Adaptive fusion of texture-based grading for Alzheimer's disease classification. Comput Med Imaging Graph. 2018;70:8–16.
15. Zheng W, Yao Z, Xie Y, Fan J, Hu B. Identification of Alzheimer's disease and mild cognitive impairment using networks constructed based on multiple morphological brain features. Biol Psychiatry Cognit Neurosci Neuroimaging. 2018;3(10):887–97. https://doi.org/10.1016/j.bpsc.2018.06.004.
16. Gupta Y, Lama RK, Kwon G-R. other: Prediction and classification of Alzheimer's disease based on combined features from apolipoprotein-e genotype, cerebrospinal fluid, MR, and FDG-PET imaging biomarkers. Front Comput Neurosci. 2019;13:72. https://doi.org/10.3389/fncom.2019.00072.
17. Aderghal K, Khvostikov A, Krylov A, Benois-Pineau J, Afdel K, Catheline G. Classification of Alzheimer disease on imaging modalities with deep CNNs using cross-modal transfer learning. In: 2018 IEEE 31st international symposium on computer-based medical systems (CBMS). 2018. p. 345–50. https://doi.org/10.1109/CBMS.2018.00067
18. Bäckström K, Nazari M, Gu IY-H, Jakola AS. An efficient 3D deep convolutional network for Alzheimer's disease diagnosis using MR images. In: 2018 IEEE 15th International symposium on biomedical imaging (ISBI 2018). 2018. p. 149–153. https://doi.org/10.1109/ISBI.2018.8363543
19. Li F, Liu M. Alzheimer's disease diagnosis based on multiple cluster dense convolutional networks. Comput Med Imaging Graph. 2018;70:101–10. https://doi.org/10.1016/j.compmedimag.2018.09.009.
20. Pan Y, Liu M, Lian C, Zhou T, Xia Y, Shen D. Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for Alzheimer's disease diagnosis. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, editors. Medical image computing and computer assisted intervention—MICCAI 2018. Cham: Springer; 2018. p. 455–63.
21. Senanayake U, Sowmya A, Dawes L. Deep fusion pipeline for mild cognitive impairment diagnosis. In: 2018 IEEE 15th International symposium on biomedical imaging (ISBI). 2018. p. 1394–1997. https://doi.org/10.1109/ISBI.2018.8363832
22. Shi J, Zheng X, Li Y, Zhang Q, Ying S. Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease. IEEE J Biomed Health Inform. 2018;22(1):173–83. https://doi.org/10.1109/JBHI.2017.2655720.
23. Stamate D, Kim M, et al. A metabolite-based machine learning approach to diagnose Alzheimer-type dementia in blood: results from the European medical information framework for Alzheimer disease biomarker discovery cohort. Alzheimer's Dement Transl Res Clin Intervent. 2019;5(1):933–8. https://doi.org/10.1016/j.trci.2019.11.001.
24. Islam J, Zhang Y. GAN-based synthetic brain PET image generation. Brain Informat. 2020;7(1):3. https://doi.org/10.1186/s40708-020-00104-2.
25. Jo T, Nho K, Risacher SL, Saykin AJ. for the Alzheimer's Neuroimaging Initiative: deep learning detection of informative features in tau PET for Alzheimer's disease classification. BMC Bioinformat. 2020;21(21):496. https://doi.org/10.1186/s12859-020-03848-0.
26. Yu W, Lei B, Ng MK, Cheung AC, Shen Y, Wang S. Tensorizing GAN with high-order pooling for Alzheimer's disease assessment. IEEE Trans Neural Netw Learn Syst. 2022;33(9):4945–59. https://doi.org/10.1109/TNNLS.2021.3063516.
27. Rahaman MM, Li C, Yao Y, Kulwa F, Rahman MA, Wang Q, Qi S, Kong F, Zhu X, Zhao X. Identification of COVID-19 samples from chest X-ray images using deep learning: a comparison of transfer learning approaches. J Xray Sci Technol. 2020;28(5):821–39. https://doi.org/10.3233/XST-200715.
28. Chen H, Li C, Wang G, Li X, Mamunur Rahaman M, Sun H, Hu W, Li Y, Liu W, Sun C, Ai S, Grzegorzek M. GasHis-Transformer: a multi-scale visual transformer approach for gastric histopathological image detection. Pattern Recogn. 2022;130: 108827. https://doi.org/10.1016/j.patcog.2022.108827.
29. Liu W, Li C, Xu N, Jiang T, Rahaman MM, Sun H, Wu X, Hu W, Chen H, Sun C, Yao Y, Grzegorzek M. CVM-Cervix: a hybrid cervical pap-smear image classification framework using CNN, visual transformer and multilayer perceptron. Pattern Recogn. 2022;130: 108829. https://doi.org/10.1016/j.patcog.2022.108829.
30. Zhang J, Li C, Kosov S, Grzegorzek M, Shirahama K, Jiang T, Sun C, Li Z, Li H. LCU-Net: a novel low-cost U-net for environmental microorganism image segmentation. Pattern Recogn. 2021;115: 107885. https://doi.org/10.1016/j.patcog.2021.107885.
31. Rahaman MM, Li C, Yao Y, Kulwa F, Wu X, Li X, Wang Q. DeepCervix: a deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques.

Comput Biol Med. 2021;136: 104649. https://doi.org/10.1016/j.compbiomed.2021.104649.
32. Guarracino MR, Maddalena L. SDI+: a novel algorithm for segmenting dermoscopic images. IEEE J Biomed Health Inform. 2019;23(22):481–8.
33. Maddalena L, Antonelli L, Albu A, Hada A, Guarracino MR. Artificial intelligence for cell segmentation, event detection, and tracking for label-free microscopy imaging. Algorithms. 2022. https://doi.org/10.3390/a15090313.
34. Aderghal K, Boissenin M, Benois-Pineau J, Catheline G, Afdel K. Classification of sMRI for AD diagnosis with convolutional neuronal networks: A pilot 2-D+ϵ study on ADNI. In: Amsaleg L et al, editors. MultiMedia Modeling—23rd International conference, MMM 2017, Reykjavik, Iceland, January 4–6, 2017, proceedings, part I. Lecture notes in computer science, Vol. 10132. Cham: Springer; 2017. p. 690–701. https://doi.org/10.1007/978-3-319-51811-4_56
35. Previtali F, Bertolazzi P, Felici G, Weitschek E. A novel method and software for automatically classifying Alzheimer's disease patients by magnetic resonance imaging analysis. Comput Methods Programs Biomed. 2017;143:89–95. https://doi.org/10.1016/j.cmpb.2017.03.006.
36. Li X, Wang H, Long J, et al. Systematic analysis and biomarker study for Alzheimer's disease. Sci Rep. 2018;8:17394. https://doi.org/10.1038/s41598-018-35789-3.
37. Voyle N, Keohane A, Newhouse S, Lunnon K, Johnston C, Soininen H, Kloszewska I, Mecocci P, Tsolaki M, Vellas B, Lovestone S, Hodges A, Kiddle S, Dobson R. A pathway based classification method for analyzing gene expression for Alzheimer's disease diagnosis. J Alzheimers Dis. 2016;49(3):659–69. https://doi.org/10.3233/JAD-150440.
38. Ding X, Bucholc M, Wang H, Glass D, Wang H, Clarke D, Bjourson A, Dowey L, O'Kane M, Prasad G, Maguire L, Wong-Lin K. A hybrid computational approach for efficient Alzheimer's disease classification based on heterogeneous data. Sci Rep. 2018. https://doi.org/10.1038/s41598-018-27997-8.
39. Nho K, et al. Integration of bioinformatics and imaging informatics for identifying rare PSEN1 variants in Alzheimer's disease. BMC Med Genom. 2016. https://doi.org/10.1186/s12920-016-0190-9.
40. Peng J, An L, Zhu X, Jin Y, Shen D. Structured sparse kernel learning for imaging genetics based Alzheimer's disease diagnosis. In: International conference on medical image computing and computer-assisted intervention. Springer; 2016. p. 70–8.
41. Maddalena L, Granata I, Manipur I, Manzo M, Guarracino M. Glioma Grade Classification Via Omics Imaging. In: Proceedings of the 13th international joint conference on biomedical engineering systems and technologies—volume 2: BIOIMAGING. Portugal: SciTePress; 2020. p. 82–92. https://doi.org/10.5220/0009167700820092. INSTICC
42. Maddalena L, Granata I, Manipur I, Manzo M, Guarracino MR. A framework based on metabolic networks and biomedical images data to discriminate glioma grades. In: Ye X, Soares F, De Maria E, Gómez Vilda P, Cabitza F, Fred A, Gamboa H, editors. Biomedical engineering systems and technologies. Springer: Cham; 2021. p. 165–89. https://doi.org/10.1007/978-3-030-72379-8_9.
43. Antonelli L, Guarracino MR, Maddalena L, Sangiovanni M. Integrating imaging and omics data: a review. Biomed Signal Process Control. 2019;52:264–80. https://doi.org/10.1016/j.bspc.2019.04.032.
44. Lovestone S, Francis P, Strandgaard K. Biomarkers for disease modification trials-the innovative medicines initiative and AddNeuroMed. J Nutr Health Aging. 2007;11(4):359–61.
45. Samper-González J, Burgos N, et al. Reproducible evaluation of classification methods in Alzheimer's disease: framework and application to MRI and PET data. Neuroimage. 2018;183:504–21. https://doi.org/10.1016/j.neuroimage.2018.08.042.
46. Routier A, Burgos N, et al. Clinica: an open-source software platform for reproducible clinical neuroscience studies. Front Neuroinformat. 2021. https://doi.org/10.3389/fninf.2021.689675.
47. Gorgolewski K, Auer T, et al. The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. Sci Data. 2016. https://doi.org/10.1038/sdata.2016.44.
48. Vapnik V. The nature of statistical learning theory. Cham: Springer; 1995.
49. Ashburner J, Friston KJ. Unified segmentation. Neuroimage. 2005;26(3):839–51. https://doi.org/10.1016/j.neuroimage.2005.02.018.
50. Ashburner J. A fast diffeomorphic image registration algorithm. Neuroimage. 2007;38(1):95–113. https://doi.org/10.1016/j.neuroimage.2007.07.007.
51. Rolls ET, Joliot M, Tzourio-Mazoyer N. Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. Neuroimage. 2015;122:1–5. https://doi.org/10.1016/j.neuroimage.2015.07.075.
52. Joliot M, Jobard G, Naveau M, Delcroix N, Petit L, Zago L, Crivello F, Mellet E, Mazoyer B, Tzourio-Mazoyer N. AICHA: an atlas of intrinsic connectivity of homotopic areas. J Neurosci Methods. 2015;254:46–59. https://doi.org/10.1016/j.jneumeth.2015.07.013.
53. Hammers A, Allom R, Koepp MJ, Free SL, Myers R, Lemieux L, Mitchell TN, Brooks DJ, Duncan JS. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Hum Brain Mapp. 2003;19(4):224–47. https://doi.org/10.1002/hbm.10123.
54. Gousias IS, Rueckert D, Heckemann RA, Dyet LE, Boardman JP, Edwards AD, Hammers A. Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest. Neuroimage. 2008;40(2):672–84. https://doi.org/10.1016/j.neuroimage.2007.11.034.
55. Shattuck DW, Mirza M, Adisetiyo V, Hojatkashani C, Salamon G, Narr KL, Poldrack RA, Bilder RM, Toga AW. Construction of a 3D probabilistic atlas of human cortical structures. Neuroimage. 2008;39(3):1064–80. https://doi.org/10.1016/j.neuroimage.2007.09.031.
56. Du P, Feng G, Kibbe W, Lin S. lumiHumanIDMapping: illumina identifier mapping for human. R package version 1.10.1. 2016.
57. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):47. https://doi.org/10.1093/nar/gkv007.
58. McCombe N, Liu S, Ding X, Prasad G, Bucholc M, Finn DP, Todd S, McClean PL, Wong-Lin K. Practical strategies for extreme missing data imputation in dementia diagnosis. IEEE J Biomed Health Informat. 2022;26(2):818–27. https://doi.org/10.1109/JBHI.2021.3098511.
59. Matthews BW. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. Biochim Biophys Acta Protein Struct. 1975;405(2):442–51. https://doi.org/10.1016/0005-2795(75)90109-9.
60. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genom. 2020. https://doi.org/10.1186/s12864-019-6413-7.
61. Lazli L, Boukadoum M, Ait Mohamed O. Computer-aided diagnosis system of Alzheimer's disease based on multimodal fusion: tissue quantification based on the hybrid fuzzy-genetic-possibilistic model and discriminative classification based on

the SVDD model. Brain Sci. 2019. https://doi.org/10.3390/brainsci9100289.
62. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. 2015. arXiv https://doi.org/10.48550/ARXIV.1511.06434.
63. Bos I, Vos S, Vandenberghe R, et al. The EMIF-AD multimodal biomarker discovery study: design, methods and cohort characteristics. Alz Res Ther. 2018. https://doi.org/10.1186/s13195-018-0396-5.
64. Kim M, Snowden S, et al. Primary fatty amides in plasma associated with brain amyloid burden, hippocampal volume, and memory in the European medical information framework for Alzheimer's disease biomarker discovery cohort. Alzheimer's Dement. 2019;15(6):817–27. https://doi.org/10.1016/j.jalz.2019.03.004.
65. Jain R, Jain N, Aggarwal A, Hemanth DJ. Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images. Cogn Syst Res. 2019;57:147–59. https://doi.org/10.1016/j.cogsys.2018.12.015.