Machine learning e modelli di classificazione di dati biomedici

Christian Salvatore
Scuola Universitaria Superiore IUSS Pavia

Decision support systems in clinical medicine

- decision support systems, designed to increase the effectiveness of the medical analysis as it provides support to clinicians who need to make strategic medical decisions in the face of a medical problems that can not be solved with operational research models

- extract from a significant amount of data, in a short time and in a versatile way, new information useful to the clinical decision-making processes

- Assisted Diagnosis
- Objective clinical assessment
- High diagnostic accuracy

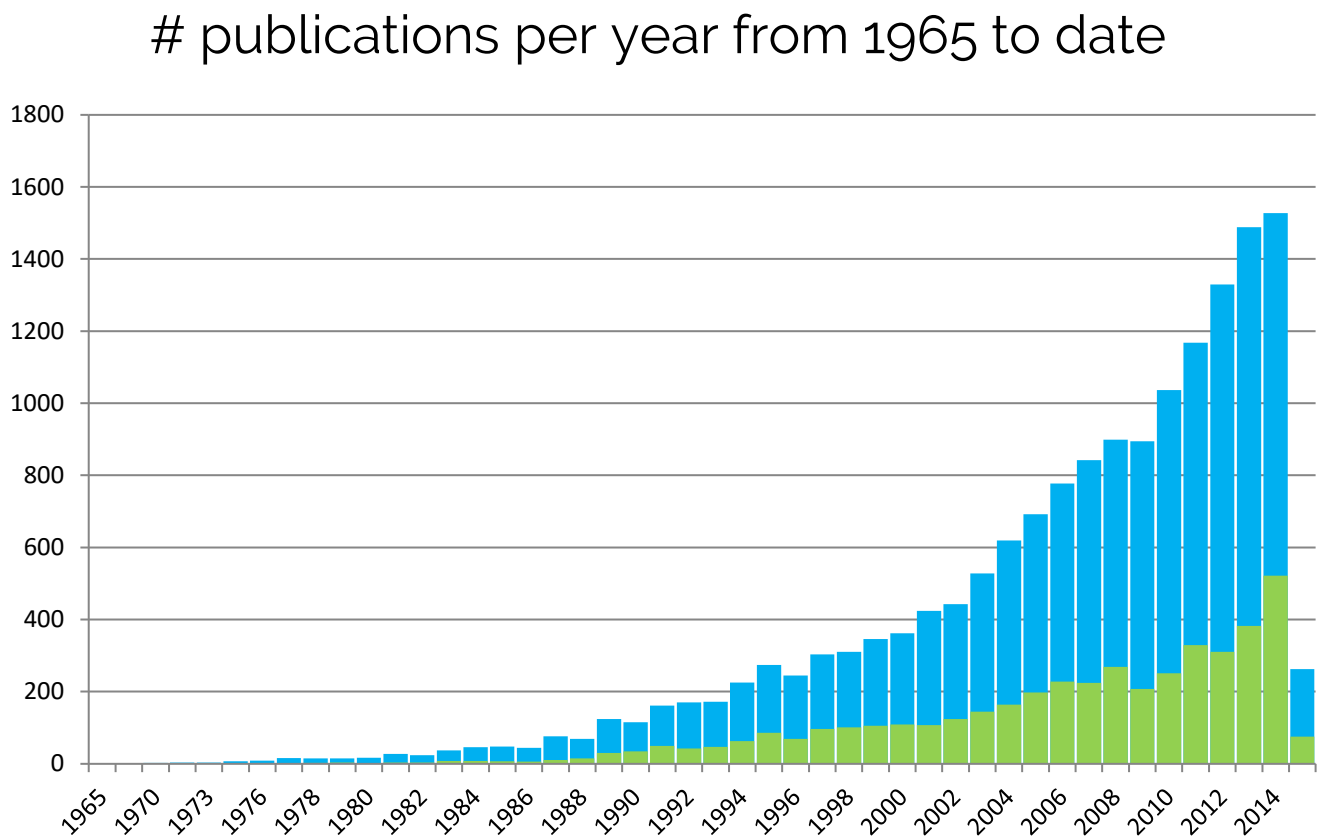
Decision support systems in clinical medicine

In vivo imaging plays a new role in the diagnostic process of many neurological diseases, including neurodegenerative diseases and psychiatric disorders

[e.g. McKhann GM 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 & Dementia 7(3): 263-69. (2011)]

Earlier and Differential Diagnosis

Decision support systems in clinical medicine



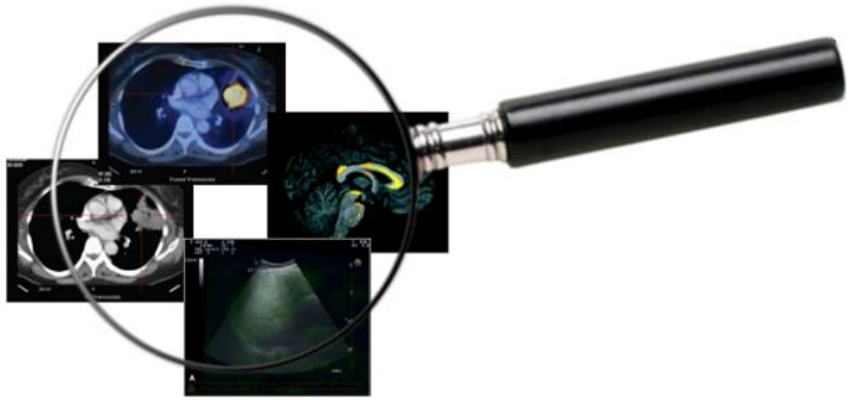
RESEARCH CRITERIA:

- “decision support system”
- “decision support system medicine”

source: pubmed.com

Machine learning applied to medical imaging

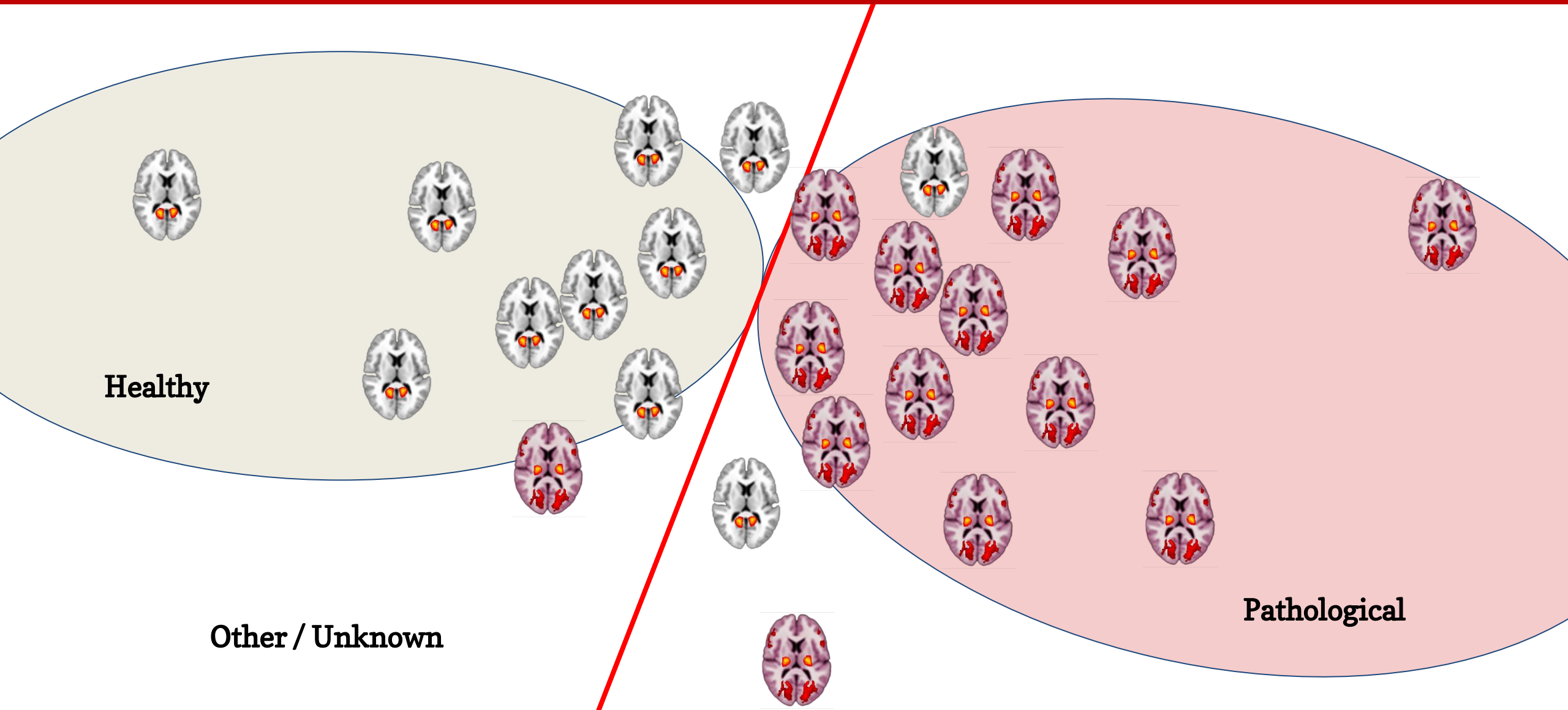
. how is this useful in medicine?



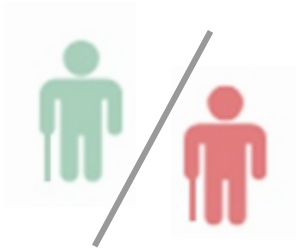
Machine learning applied to medical imaging

Machine learning applied to medical imaging

Machine learning applied to medical imaging



Machine learning applied to medical imaging



Diagnosis (early/differential)

is the patient healthy?

Prognosis

what will be the course of the disease?

Screening

are the patients -within a population- healthy?

Treatment addressing

will this therapy be effective for this patient?

AI Applications to Medical Imaging

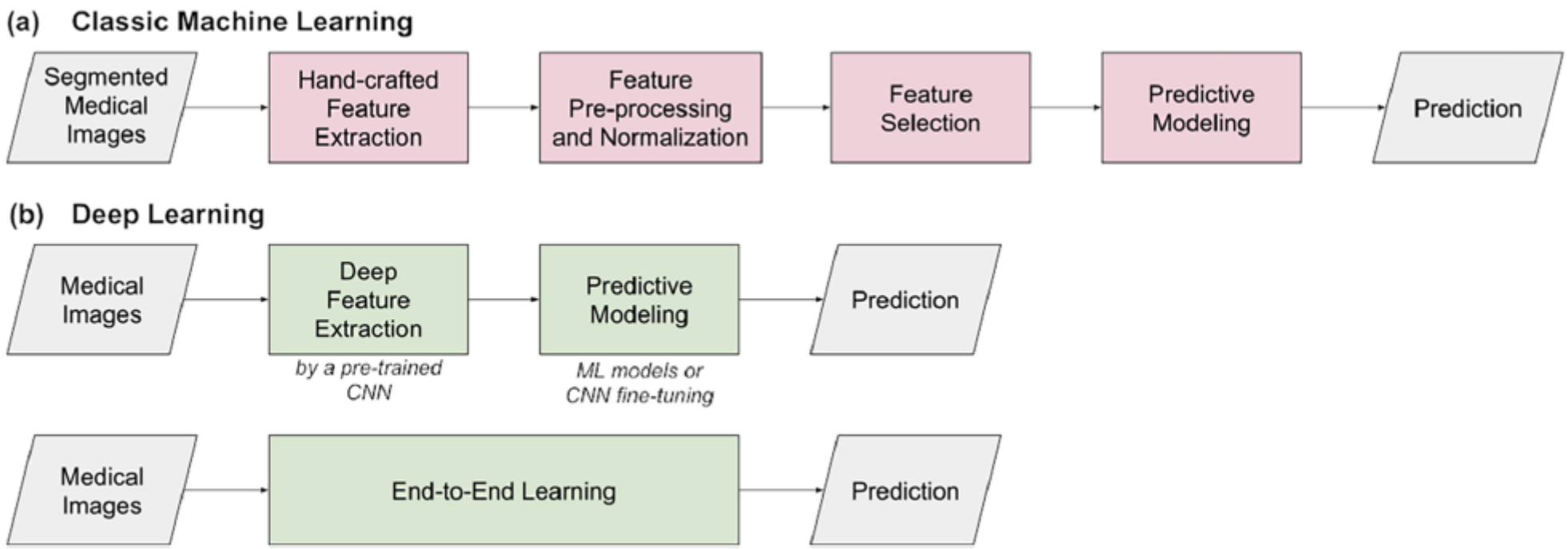
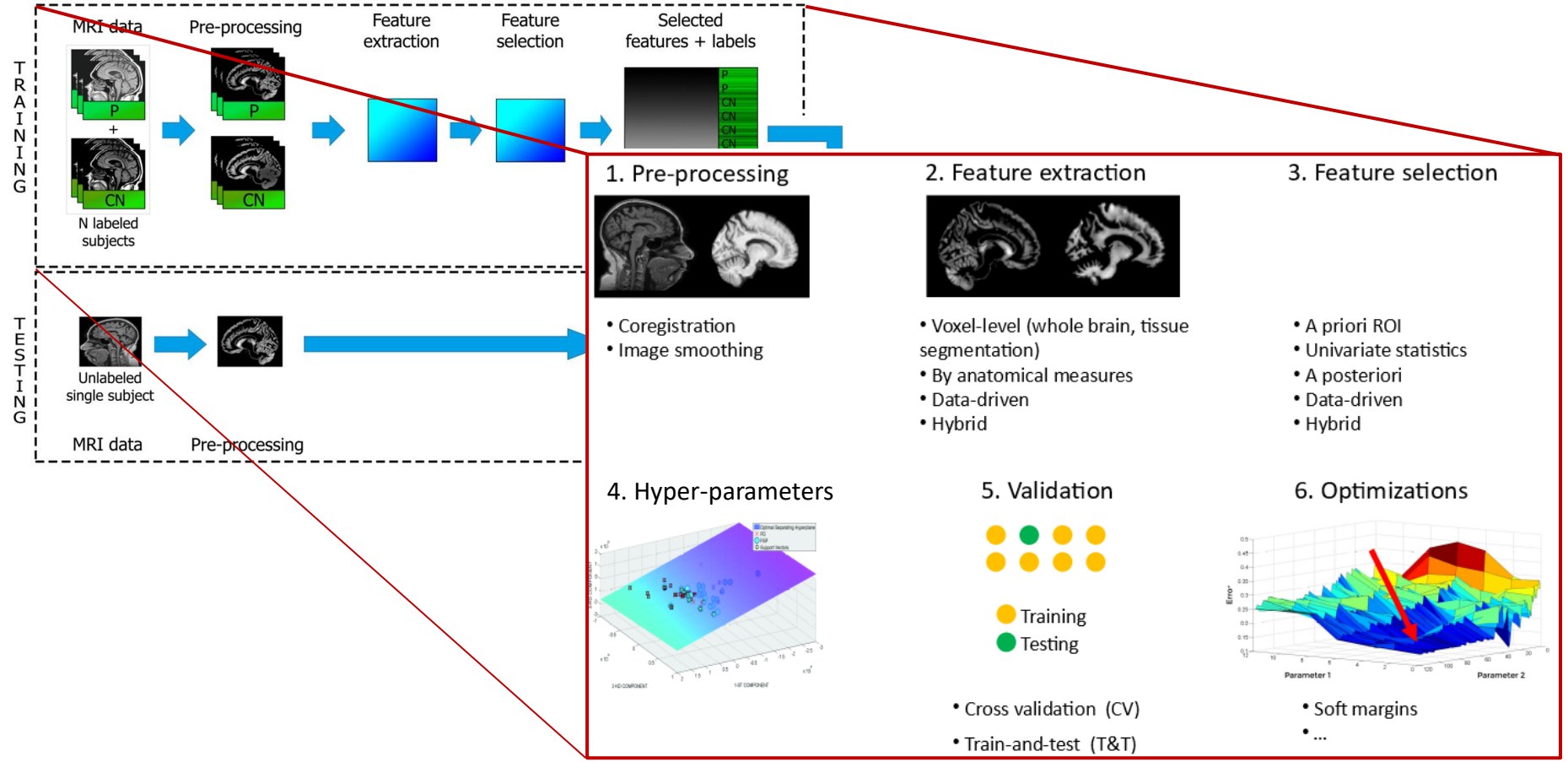
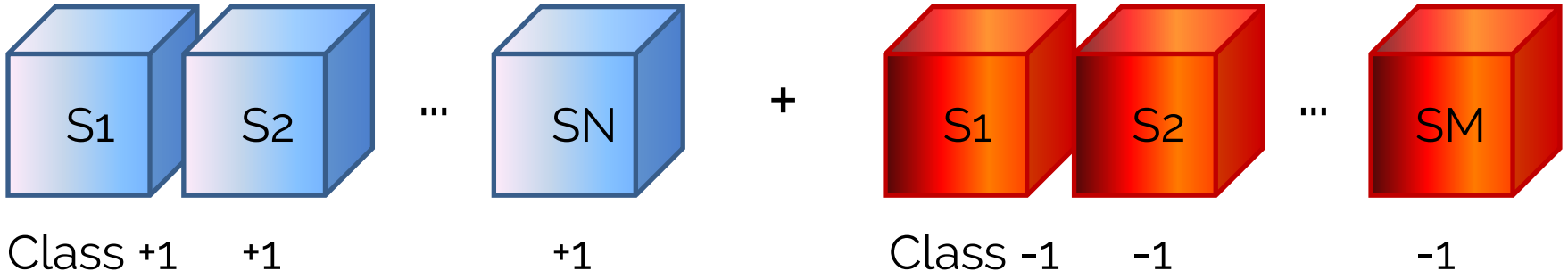


Fig. 1. Typical architecture and workflow of artificial intelligence systems for predictive modelling: a) classic machine learning, with the various processing steps involving hand-crafted features such as in radiomics; b) deep learning considering either deep medical image feature extraction or end-to-end learning.

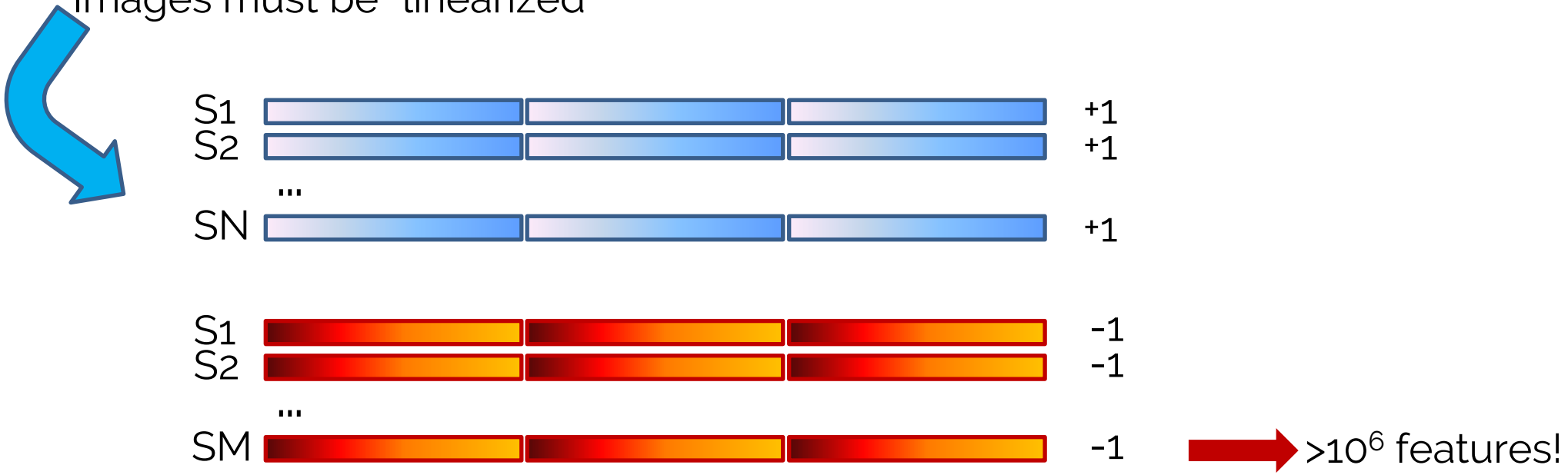
AI Applications to Medical Imaging



AI Applications to Medical Imaging



Images must be “linearized”



AI Applications to Medical Imaging

Supervised OR Unsupervised Learning?

IS A CLINICAL ENDPOINT AVAILABLE?

Training / Validation / Testing

GENERALIZATION ABILITY?
WHICH PROPORTION OF DATA FOR TESTING?
WHICH VALIDATION APPROACH?

Data quantity – How many?

SAMPLE SIZE

DATA AUGMENTATION: DATA WARPING, OVERSAMPLING, GAN

IMBALANCE LEARNING: RESAMPLING

ENSEMBLE LEARNING

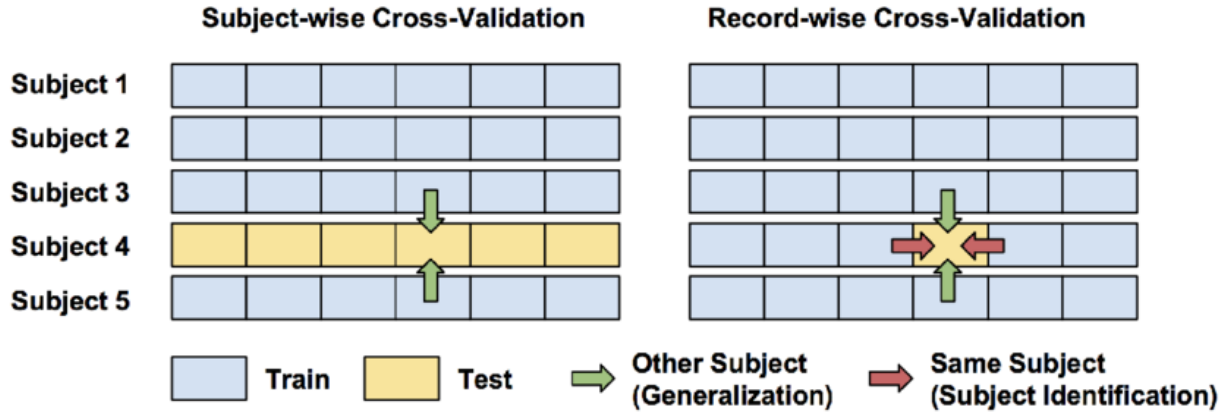
CAUSES OF BIAS?

Causes of bias

Record-Wise Validation

Record-wise vs. subject-wise validation

"Record-wise cross validation typically inflates the prediction accuracy, and subject-wise cross validation is a more desired and appropriate way of evaluating the performance of automatic classification"



Circularity

more likely to develop dementia with Lewy body while patients with aMCI are more likely to convert to AD [17].

This study investigates the classification of AD, aMCI, and naMCI by combining subcortical volumes of MRI with a neuropsychological test (mini-mental state examination (MMSE)), which is most often administered to screen patients for cognitive impairment and dementia. This study demonstrates the merits of MMSE and extends its use to the discrimination of different stages of AD when used in conjunction with select volumetric variables. To the best of our knowledge, this study is the first that investigates the impact of combining MRI at baseline with MMSE for the detection of AD, aMCI, and naMCI using support vector machine (SVM) methodology. Another important contribution of this study is the development of a fully automated feature extraction technique, which in its initial step associates equal weights to each of the measured volumes, and yet as its outcome is a ranking of the volumes that can be used as variables in a multidimensional decisional space for optimal classification.

II. Methodology

The general structure of the proposed approach is presented in Fig. 1 showing the main steps of the whole process from acquisition of the MRI scans, through the sorting and selection of variables or features that will constitute the decisional space for the classification process using the well-established SVM classifier. The proposed approach is also open to the use of other alternative classification algorithms such as artificial neural networks, optimal discriminant analysis, and so on. This study opted for SVM only for its implementation simplicity.

A. Subjects

In this study, a total of 309 participants were recruited from Wien Center for Alzheimer's Disease and Memory Disorders, Mount Sinai Medical Center, Miami Beach, FL as shown in Table I between 2005 and 2008. All participants have taken the Folstein MMSE [18] with a minimum score of 15. The study

Fig. 1. General structure of the classification approach.

was approved by the Mount Sinai Medical Center Institutional Review Board with informed consent provided by the subjects or legal representatives.

All subjects had: 1) a neurological and medical evaluation by a physician; 2) MMSE; 3) a structural volumetrically acquired MRI scan of the brain. MMSE was used as the index of cognitive ability and sum of boxes from the Clinical Dementia Rating Scale (CDR-sb) was used clinically as the index of functional ability.

The cognitive diagnosis was made using a combination of the physician's diagnosis and the neuropsychological diagnosis, as described previously [19]. The etiological diagnosis was made by the examining physician. The diagnosis of cognitive normal (CN) required that the physician's diagnosis was CN and no cognitive test scores were $\geq$1.5 S.D. below age and education-corrected means. A probable AD diagnosis required a dementia syndrome and the National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer's Disease and Related Disorders Association criteria for AD [20].

The diagnosis of aMCI was rendered by a clinical impression by the examining physician of a history of MCI but no significant functional impairment and did not meet Diagnostic and Statistical Manual of Mental Disorder-4th edition (DSM-IV) criteria [21] for dementia. This diagnosis was confirmed by a neuropsychological evaluation in which one or more *tests of memory* had to fall 1.5 S.D. or more below expected normative values.

The diagnosis of naMCI was rendered by a clinical impression by the examining physician of a history of MCI but no significant functional impairment and did not meet DSM-IV criteria for dementia. This diagnosis was confirmed by a neuropsychological evaluation in which one or more tests of *nonmemory function* (e.g., Trails B, Similarities, and Category Fluency) had

Circularity

more likely to develop dementia with Lewy body while patients with aMCI are more likely to convert to AD [17].

This study investigates the classification of AD, aMCI, and naMCI by combining subcortical volumes of MRI with a neuropsychological test (mini-mental state examination (MMSE)), which is most often administered to screen patients for cognitive impairment and dementia. This study demonstrates the merits of MMSE and extends its use to the discrimination of different stages of AD when used in conjunction with select volumetric variables. To the best of our knowledge, this study is the first that investigates the impact of combining MRI at baseline with MMSE for the detection of AD, aMCI, and naMCI using support vector machine (SVM) methodology. Another important contribution of this study is the development of a fully automated feature extraction technique, which in its initial step associates equal weights to each of the measured volumes, and yet as its outcome is a ranking of the volumes that can be used as variables in a multidimensional decisional space for optimal classification.

II. Methodology

The general structure of the proposed approach is presented in Fig. 1 showing the main steps of the whole process from acquisition of the MRI scans, through the sorting and selection of variables or features that will constitute the decisional space for the classification process using the well-established SVM classifier. The proposed approach is also open to the use of other alternative classification algorithms such as artificial neural networks, optimal discriminant analysis, and so on. This study opted for SVM only for its implementation simplicity.

A. Subjects

In this study, a total of 309 participants were recruited from Wien Center for Alzheimer's Disease and Memory Disorders, Mount Sinai Medical Center, Miami Beach, FL as shown in Table I between 2005 and 2008. All participants have taken the Folstein MMSE [18] with a minimum score of 15. The study

Fig. 1. General structure of the classification approa

was approved by the Mount Sinai Medical
Review Board with informed consent prov
or legal representatives.

All subjects had: 1) a neurological and m
a physician; 2) MMSE; 3) a structural volu
MRI scan of the brain. MMSE was used as th
ability and sum of boxes from the Clinic
Scale (CDR-sb) was used clinically as the
ability.

The cognitive diagnosis was made usin
the physician's diagnosis and the neurops
sis, as described previously [19]. The etiolo
made by the examining physician. The dia
normal (CN) required that the physician's
and no cognitive test scores were ≥1.5 S
education-corrected means. A probable AD
a dementia syndrome and the National Insti
and Communicative Disorders and Stroke/A
and Related Disorders Association criteria

The diagnosis of aMCI was rendered b
sion by the examining physician of a hist
significant functional impairment and did n
and Statistical Manual of Mental Disorder
IV) criteria [21] for dementia. This diagnos
a neuropsychological evaluation in which
memory had to fall 1.5 S.D. or more below
values.

The diagnosis of naMCI was rendered by
by the examining physician of a history of
icant functional impairment and did not me
for dementia. This diagnosis was confirmed
logical evaluation in which one or more t
function (e.g., Trails B, Similarities, and Ca

D spoiled gradient echo sequences (FSPGR). Specifications for 3-D MPRAGE include coronal sections with a 1.5-mm gap in thickness; section interval, 0.75 mm; TR, 2190 ms; TE, 4.38 ms; TI, 1100 ms; FA, 15°; NEX, 1; matrix, 256 × 256; FOV, 260 mm; bandwidth, 130 Hz/pixel; acquisition time, 9 min.; phase-encoding direction, right to left. Specifications for 3-D FSPGR were the following: 140 contiguous coronal sections of 1.2-mm thickness; contiguous images with no section interval; TR, 7.8 ms; TE, 3.0 ms; inversion recovery preparation time, 450 ms; flip angle, 12°; NEX, 1; matrix, 256 × 256; FOV, 240 mm; bandwidth, 31.25 Hz/pixel; acquisition time, 6–7 min.; phase-encoding direction, right to left.

C. Image Analysis

FreeSurfer pipeline (version 5.1.0) was applied to the MRI scans to produce 55 volumetric variables, including 45 subcortical regions (e.g., left lateral ventricle, corpus callosum anterior, right hippocampus, etc.) and 10 morphometric statistics (e.g., left hemisphere gray matter volume, total cortical volume, etc.). Out of the 45 volumetric variables, four of them, namely left white-matter-hypointensities (WMH), right WMH, left non-WMH, and right non-WMH were excluded since they were all characterized by zero values. Therefore, each MRI scan includes 41 regional and 10 morphometric volumes. It was determined that MRI scans from the two scanner machines did not change the variance of volume difference when comparing subcortical volumes (FreeSurfer segmentation) from the test–retest scans acquired in a fixed machine [22], [23], thus no correction is needed for scanner difference.

D. Feature Extraction and Statistical Significance

AD patients suffer from cerebral atrophy, which can be distinguished from normal aging [3], and specific regions are more atrophied along the progression of AD. For example, studies have shown that hippocampal atrophy is more significant as disease progresses [24]. Determination of the key atrophied/enlarged

All volumetric variables, but for intracranial (ICV), were adjusted for ICV, age, and education as per (1), as they were found to be significant factors as demonstrated in Table I:

$$V_a = V_{ua} - G_{\mathrm{ICV}} \cdot (V_{s\mathrm{ICV}} - V_{m\mathrm{ICV}}) - G_{\mathrm{EDU}} \cdot (E_s - E_m) - G_{\mathrm{AGE}} \cdot (A_s - A_m) \quad (1)$$

where V_a is the adjusted volume, V_{ua} is the unadjusted volume, $V_{s\mathrm{ICV}}$, E_s, and A_s are the subject ICV, years of education, and age (years), respectively; $V_{m\mathrm{ICV}}$, E_m, and A_m are the corresponding means for all the control subjects. The gradients G_{ICV}, G_{EDU}, and G_{AGE} were derived by a region specific regression against subject ICV, years of education, and age of all the participants so that the regression is fully blinded to the classifications. As per Chiang *et al.* [25], the above regression also has the advantage that the regressing order of the three factors does not affect the results.

The adjusted volumes and ICV of the 51 volumetric variables are then combined with the MMSE score to generate a 52-variable vector discriminator for each subject. A Student's t-test is carried out on each of the 52 variables between AD (or MCI) and CN to determine the significance of each variable in the classification outcome and only those with a p-value lower than significance level (α) of 0.05 are selected and ranked.

It should also be noted that even though atrophy is what is generally sought, statistical testing in this study considers both cases of atrophy and enlargement of brain regions, since volumetric enlargement (i.e., ventricles filled with cerebrospinal fluid) is also shown to be an important predictor of AD [26].

E. Variable Selection Using Incremental Error Analysis

Rank of the statistically significant variables provides an overall view of the discriminative power of each variable for each classification type. Selection of these optimal variables can be viewed as a dimensionality reduction problem, which is performed using an incremental error analysis. The result of this analysis is the determination of how many of these top-ranked

Circularity

Circularity / Double dipping

Data used for training the classifier or for optimizing the parameters of the model (including feature extraction/selection or hyperparameter tuning) are the same used for testing the generalization ability

Circular analysis in systems neuroscience – the dangers of double dipping

Nikolaus Kriegeskorte, **W Kyle Simmons**, **Patrick SF Bellgowan**, and **Chris I Baker**
Laboratory of Brain and Cognition, National Institute of Mental Health

Abstract

A neuroscientific experiment typically generates a large amount of data, of which only a small fraction is analyzed in detail and presented in a publication. However, selection among noisy measurements can render circular an otherwise appropriate analysis and invalidate results. Here we argue that systems neuroscience needs to adjust some widespread practices in order to avoid the circularity that can arise from selection. In particular, "double dipping" – the use of the same data set for selection and selective analysis – will give distorted descriptive statistics and invalid statistical inference whenever the results statistics are not inherently independent of the selection criteria under the null hypothesis. To demonstrate the problem, we apply widely used analyses to noise data known not to contain the experimental effects in question. Spurious effects can appear in the context of both univariate activation analysis and multivariate pattern-information analysis. We suggest a policy for avoiding circularity.

Although the dangers of double dipping in the pool of data are well understood in statistics and computer science, the practice is common in systems neuroscience, and in particular in neuroimaging and electrophysiology. To assess how widespread nonindependent selective analyses are in the literature, we examined all functional-magnetic-resonance-imaging (fMRI) studies published in five prestigious journals (Nature, Science, Nature Neuroscience, Neuron, Journal of Neuroscience) in 2008. Of these 134 fMRI papers, 42% (57 papers) contained at least one nonindependent selective analysis (not considering supplementary materials). Another 14% (20 papers) may contain nonindependent selective analyses, but the methodological information given was insufficient to reach a judgment.

Circularity

Possible solutions

Make the gold standard completely independent from the input data

Hindawi Publishing Corporation
Behavioural Neurology
Volume 2017, Article ID 1850909, 19 pages
https://doi.org/10.1155/2017/1850909

Hindawi

Research Article

Optimizing Neuropsychological Assessments for Cognitive, Behavioral, and Functional Impairment Classification: A Machine Learning Study

Petronilla Battista, Christian Salvatore, and Isabella Castiglioni

Institute of Molecular Bioimaging and Physiology, National Research Council (IBFM-CNR), Segrate, Milano, Italy

Correspondence should be addressed to Isabella Castiglioni; isabella.castiglioni@ibfm.cnr.it

Received 27 May 2016; Revised 7 December 2016; Accepted 21 December 2016; Published 31 January 2017

Academic Editor: Michael Nitsche

Copyright © 2017 Petronilla Battista et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Subjects with Alzheimer's disease (AD) show loss of cognitive functions and change in behavioral and functional state affecting the quality of their daily life and that of their families and caregivers. A neuropsychological assessment plays a crucial role in detecting such changes from normal conditions. However, despite the existence of clinical measures that are used to classify and diagnose AD, a large amount of subjectivity continues to exist. Our aim was to assess the potential of machine learning in quantifying this process and optimizing or even reducing the amount of neuropsychological tests used to classify AD patients, also at an early stage of impairment. We investigated the role of twelve state-of-the-art neuropsychological tests in the automatic classification of subjects with none, mild, or severe impairment as measured by the clinical dementia rating (CDR). Data were obtained from the ADNI database. In the groups of measures used as features, we included measures of both cognitive domains and subdomains. Our findings show that some tests are more frequently best predictors for the automatic classification, namely, LM, ADAS-Cog, AVLT, and FAQ, with a major role of the ADAS-Cog measures of delayed and immediate memory and the FAQ measure of financial competency.

1. Introduction

Dementia is a clinical syndrome which affected more than 35 million people worldwide in 2010, with new estimates of 48.1 million people for 2020 and numbers expected to almost double every 20 years [1]. Alzheimer's disease (AD) represents the primary cause of neurodegenerative dementia [2].

To date, scientists have concentrated on untangling the complex brain changes involved in the onset and progression of AD. However, this pathology is correlated to cognitive impairment, behavioral disturbance, and functional disabilities, which greatly have an impact on the quality of daily life, and is major problem for families, caregivers, and healthcare institutions. It is thus crucial to detect such changes early and to identify the level and the type of impairment in the patients. This could facilitate the provision of optimal support as soon as possible, in order to maintain their quality of life for as long as possible. In addition, early detection enables the disease to be monitored from its initial stage of disability, possibly administering available treatments when loss of functions is not yet advanced.

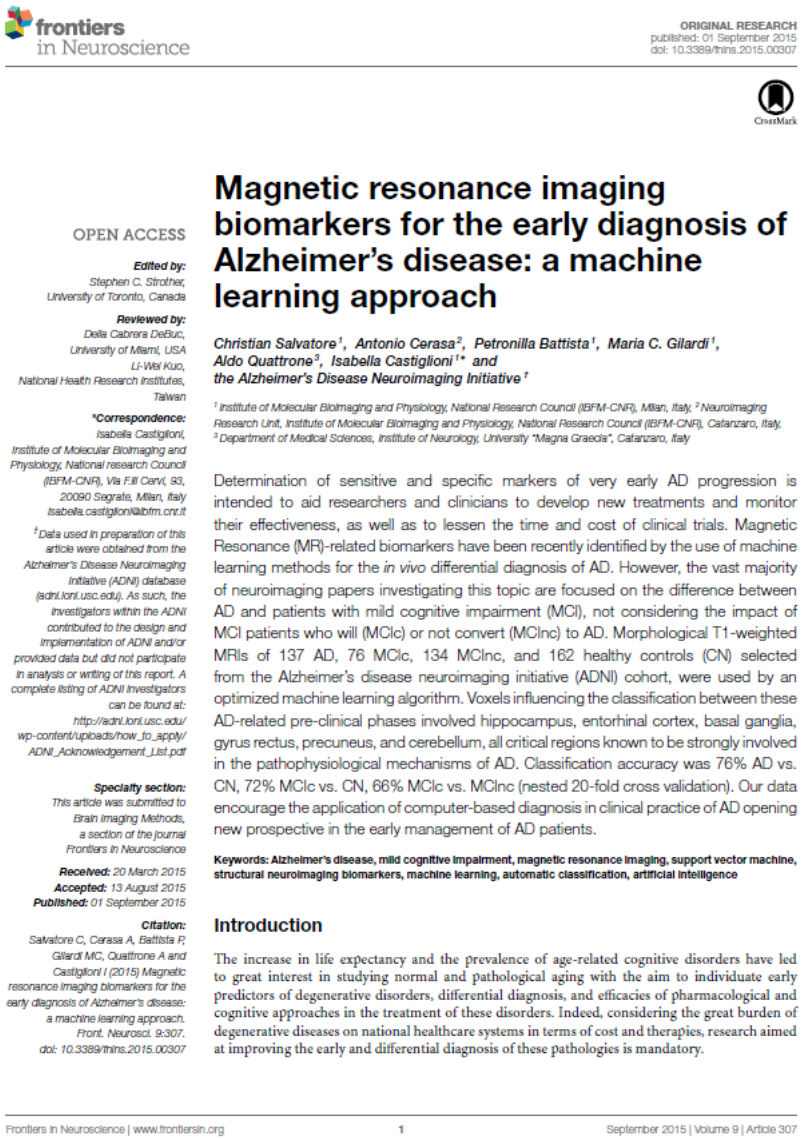
Neuropsychological assessment plays a crucial role in detecting loss of cognitive functions and change in behavioral and functional state from normal conditions. Specifically, neuropsychological tests can detect dysfunctions in human "cognitive domains" as a consequence of dysfunctions in different neural networks and subnetworks caused by AD. In 2013, the American Psychiatric Association published the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [3]. DSM-5 defined six key domains of cognitive function, namely, complex attention, executive function, learning and memory, language, perceptual-motor function, and social cognition, and each of these has subdomains. Identifying the domains and subdomains affected in a patient helps in establishing the aetiology and severity of the neurocognitive disorder. Neuropsychological tests can measure different cognitive domains (e.g., language, learning, and memory) and subdomains (e.g., long-term memory and

Circularity

Possible solutions

Make the gold standard completely independent from the input data

Measure the gold standard (e.g. diagnosis) at a follow-up date



frontiers in Neuroscience

ORIGINAL RESEARCH
published: 01 September 2015
doi: 10.3389/fnins.2015.00307

CrossMark

Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach

Christian Salvatore[1], *Antonio Cerasa*[2], *Petronilla Battista*[1], *Maria C. Gilardi*[1], *Aldo Quattrone*[3], *Isabella Castiglioni*[1]* and *the Alzheimer's Disease Neuroimaging Initiative*[†]

[1] Institute of Molecular Bioimaging and Physiology, National Research Council (IBFM-CNR), Milan, Italy; [2] Neuroimaging Research Unit, Institute of Molecular Bioimaging and Physiology, National Research Council (IBFM-CNR), Catanzaro, Italy; [3] Department of Medical Sciences, Institute of Neurology, University "Magna Graecia", Catanzaro, Italy

Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as to lessen the time and cost of clinical trials. Magnetic Resonance (MR)-related biomarkers have been recently identified by the use of machine learning methods for the *in vivo* differential diagnosis of AD. However, the vast majority of neuroimaging papers investigating this topic are focused on the difference between AD and patients with mild cognitive impairment (MCI), not considering the impact of MCI patients who will (MCIc) or not convert (MCInc) to AD. Morphological T1-weighted MRIs of 137 AD, 76 MCIc, 134 MCInc, and 162 healthy controls (CN) selected from the Alzheimer's disease neuroimaging initiative (ADNI) cohort, were used by an optimized machine learning algorithm. Voxels influencing the classification between these AD-related pre-clinical phases involved hippocampus, entorhinal cortex, basal ganglia, gyrus rectus, precuneus, and cerebellum, all critical regions known to be strongly involved in the pathophysiological mechanisms of AD. Classification accuracy was 76% AD vs. CN, 72% MCIc vs. CN, 66% MCIc vs. MCInc (nested 20-fold cross validation). Our data encourage the application of computer-based diagnosis in clinical practice of AD opening new prospective in the early management of AD patients.

Keywords: Alzheimer's disease, mild cognitive impairment, magnetic resonance imaging, support vector machine, structural neuroimaging biomarkers, machine learning, automatic classification, artificial intelligence

Introduction

The increase in life expectancy and the prevalence of age-related cognitive disorders have led to great interest in studying normal and pathological aging with the aim to individuate early predictors of degenerative disorders, differential diagnosis, and efficacies of pharmacological and cognitive approaches in the treatment of these disorders. Indeed, considering the great burden of degenerative diseases on national healthcare systems in terms of cost and therapies, research aimed at improving the early and differential diagnosis of these pathologies is mandatory.

OPEN ACCESS

Edited by:
Stephen C. Strother, University of Toronto, Canada

Reviewed by:
Della Cabrera DeBuc, University of Miami, USA
Li-Wei Kuo, National Health Research Institutes, Taiwan

***Correspondence:**
Isabella Castiglioni, Institute of Molecular Bioimaging and Physiology, National research Council (IBFM-CNR), Via Fil Cervi, 93, 20090 Segrate, Milan, Italy
isabella.castiglioni@ibfm.cnr.it

[†] *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*

Specialty section:
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroscience

Received: *20 March 2015*
Accepted: *13 August 2015*
Published: *01 September 2015*

Citation:
Salvatore C, Cerasa A, Battista P, Gilardi MC, Quattrone A and Castiglioni I (2015) Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: a machine learning approach. Front. Neurosci. 9:307. doi: 10.3389/fnins.2015.00307

Stability of diagnosis

NeuroImage 56 (2011) 766–781

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Supplement

Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database

Rémi Cuingnet [a,b,c,d,*], Emilie Gerardin [a,b,c], Jérôme Tessieras [a,b,c], Guillaume Auzias [a,b,c], Stéphane Lehéricy [a,b,c,e], Marie-Odile Habert [d,f], Marie Chupin [a,b,c], Habib Benali [d], Olivier Colliot [a,b,c] and The Alzheimer's Disease Neuroimaging Initiative [1]

[a] UPMC Université Paris 6, UMR 7225, UMR_S 975, Centre de Recherche de l'Institut du Cerveau et de la Moelle épinière (CRICM), Paris, F-75013, France
[b] CNRS, UMR 7225, CRICM, Paris, F-75013, France
[c] Inserm, UMR_S 975, CRICM, Paris, F-75013, France
[d] Inserm, UMR_S 678, LIF, Paris, F-75013, France
[e] Centre for Neuroimaging Research, CENIR, Department of Neuroradiology, Groupe hospitalier Pitié-Salpêtrière, Paris, F-75013, France
[f] AP-HP, Department of Nuclear Medicine, Groupe hospitalier Pitié-Salpêtrière, Paris, F-75013, France

ARTICLE INFO

Article history:
Received 27 November 2009
Revised 31 May 2010
Accepted 5 June 2010
Available online 11 June 2010

Keywords:
Alzheimer's disease
AD
MCI
Converter
Prodromal
Classification
Magnetic resonance imaging
Support vector machines

ABSTRACT

Recently, several high dimensional classification methods have been proposed to automatically discriminate between patients with Alzheimer's disease (AD) or mild cognitive impairment (MCI) and elderly controls (CN) based on T1-weighted MRI. However, these methods were assessed on different populations, making it difficult to compare their performance. In this paper, we evaluated the performance of ten approaches (five voxel-based methods, three methods based on cortical thickness and two methods based on the hippocampus) using 509 subjects from the ADNI database. Three classification experiments were performed: CN *vs* AD, CN *vs* MCIc (MCI who had converted to AD within 18 months, MCI converters – MCIc) and MCIc *vs* MCInc (MCI who had not converted to AD within 18 months, MCI non-converters – MCInc). Data from 81 CN, 67 MCInc, 39 MCIc and 69 AD were used for training and hyperparameters optimization. The remaining independent samples of 81 CN, 67 MCInc, 37 MCIc and 68 AD were used to obtain an unbiased estimate of the performance of the methods. For AD *vs* CN, whole-brain methods (voxel-based or cortical thickness-based) achieved high accuracies (up to 81% sensitivity and 95% specificity). For the detection of prodromal AD (CN *vs* MCIc), the sensitivity was substantially lower. For the prediction of conversion, no classifier obtained significantly better results than chance. We also compared the results obtained using the DARTEL registration to that using SPM5 unified segmentation. DARTEL significantly improved six out of 20 classification experiments and led to lower results in only two cases. Overall, the use of feature selection did not improve the performance but substantially increased the computation times.

© 2010 Elsevier Inc. All rights reserved.

Introduction

Alzheimer's disease (AD) is the most frequent neurodegenerative dementia and a growing health problem. Definite diagnosis can only be made postmortem, and requires histopathological confirmation of amyloid plaques and neurofibrillary tangles. Early and accurate diagnosis of Alzheimer's Disease (AD) is not only challenging, but is crucial in the perspective of future treatments. Clinical diagnostic criteria are currently based on the clinical examination and neuropsychological assessment, with the identification of dementia and then of the Alzheimer's phenotype (Blennow et al., 2006). Patients suffering from AD at a prodromal stage are, mostly, clinically classified as amnestic mild cognitive impairment (MCI) (Petersen et al., 1999; Dubois and Albert, 2004), but not all patients with amnestic MCI will develop AD. Recently, more precise research criteria were proposed for the early diagnostic of AD at the prodromal stage of the disease (Dubois et al., 2007). These criteria are based on a clinical core of early episodic memory impairment and the presence of at least one additional supportive feature including abnormal MRI and PET

* Corresponding author. CRICM, Equipe Cogimage (ex LENA), Hôpital de la Pitié-Salpêtrière, 47, boulevard de l'Hôpital, 75651 Paris Cedex 13, France.
E-mail address: remi.cuingnet@gmail.com (R. Cuingnet).

[1] Data used in the preparation of this article were obtained from the Alzheimer's

frontiers in Aging Neuroscience

ORIGINAL RESEARCH
published: 24 May 2018
doi: 10.3389/fnagi.2018.00135

MRI Characterizes the Progressive Course of AD and Predicts Conversion to Alzheimer's Dementia 24 Months Before Probable Diagnosis

Christian Salvatore[1], Antonio Cerasa[2] and Isabella Castiglioni[1] for the Alzheimer's Disease Neuroimaging Initiative[†]*

[1] Institute of Molecular Bioimaging and Physiology, National Research Council (IBFM-CNR), Milan, Italy, [2] Institute of Molecular Bioimaging and Physiology, National Research Council (IBFM-CNR), Catanzaro, Italy

OPEN ACCESS

Edited by:
Juan Manuel Gorriz,
Universidad de Granada, Spain

Reviewed by:
Li Su,
University of Cambridge, United Kingdom
Guido Gainotti,
Università Cattolica del Sacro Cuore, Italy

**Correspondence:*
Isabella Castiglioni
Isabella.castiglioni@ibfm.cnr.it

[†]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

Received: 15 December 2017
Accepted: 23 April 2018
Published: 24 May 2018

Citation:
Salvatore C, Cerasa A and Castiglioni I (2018) MRI Characterizes the Progressive Course of AD and Predicts Conversion to Alzheimer's Dementia 24 Months Before Probable Diagnosis. Front. Aging Neurosci. 10:135. doi: 10.3389/fnagi.2018.00135

There is no disease-modifying treatment currently available for AD, one of the more impacting neurodegenerative diseases affecting more than 47.5 million people worldwide. The definition of new approaches for the design of proper clinical trials is highly demanded in order to achieve non-confounding results and assess more effective treatment. In this study, a cohort of 200 subjects was obtained from the Alzheimer's Disease Neuroimaging Initiative. Subjects were followed-up for 24 months, and classified as AD (50), progressive-MCI to AD (50), stable-MCI (50), and cognitively normal (50). Structural T1-weighted MRI brain studies and neuropsychological measures of these subjects were used to train and optimize an artificial-intelligence classifier to distinguish mild-AD patients who need treatment (AD + pMCI) from subjects who do not need treatment (sMCI + CN). The classifier was able to distinguish between the two groups 24 months before AD definite diagnosis using a combination of MRI brain studies and specific neuropsychological measures, with 85% accuracy, 83% sensitivity, and 87% specificity. The combined-approach model outperformed the classification using MRI data alone (72% classification accuracy, 69% sensitivity, and 75% specificity). The patterns of morphological abnormalities localized in the temporal pole and medial-temporal cortex might be considered as biomarkers of clinical progression and evolution. These regions can be already observed 24 months before AD definite diagnosis. The best neuropsychological predictors mainly included measures of functional abilities, memory and learning, working memory, language, visuoconstructional reasoning, and complex attention, with a particular focus on some of the sub-scores of the FAQ and AVLT tests.

Keywords: artificial intelligence, Alzheimer's disease, clinical trials, magnetic resonance imaging, neuropsychological tests, biomarkers, predictors

INTRODUCTION

According to the World Health Organization, there were 47.5 million people worldwide with dementia in 2015, with 7.7 million new cases each year. The total number of people with dementia is projected to reach 75.6 millions in 2030 and almost triple by 2050 to 135.5 millions (Dementia Statistics, 2015; World Alzheimer Report, 2015; Khan et al., 2017). The most frequent dementia

AI Applications to Medical Imaging

Data Curation

DATA LABELLING / ANNOTATION

DATA HARMONIZATION

IMAGE-INTENSITY NORMALIZATION

DENOISING

ARTIFACT CORRECTION

New Frontiers: Federated Learning

AI Applications to Medical Imaging

Interpretability

EXPLAINABLE AI

SURROGATE BIOMARKERS

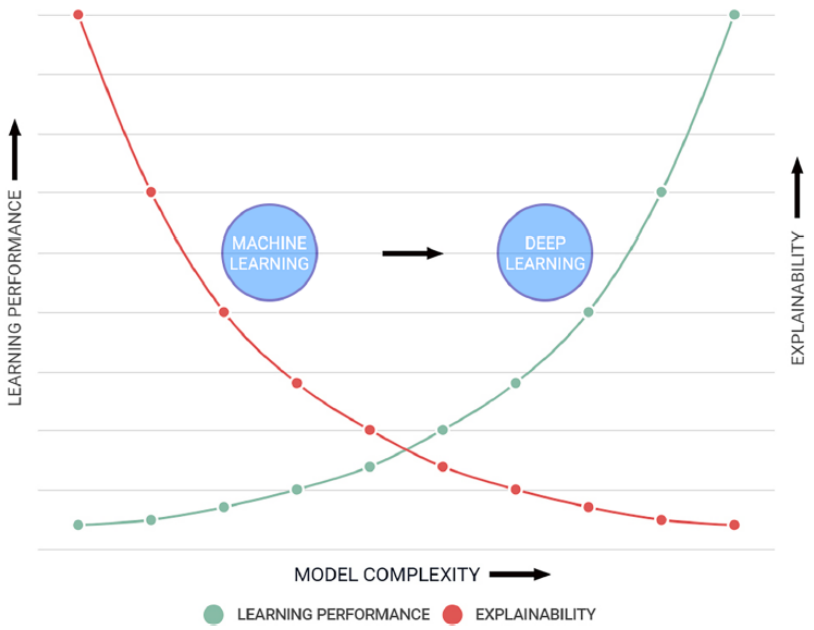


Fig. 2. Learning performance and explainability of an artificial intelligence system as a function of model complexity.

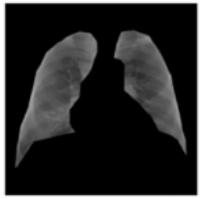
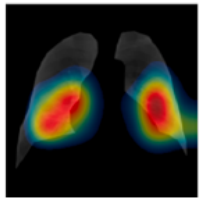
Task	Input data	AI technique	AI output	XAI output
Lesion Classification	Clinical features + Imaging features	Machine Learning (Feature selection + SVM classification)	Classification label (Malignant vs. Benign)	Classification label + Most important features for AI model: • Lesion heterogeneity • Lesion entropy • Family history
Pneumonia Diagnosis	X-Ray Imaging 	Deep Learning (Convolutional Neural Networks)	Classification label (Pneumonia vs. Healthy)	Classification label + Activation map 

Fig. 4. Representative examples of artificial intelligence (AI) tasks in medicine and corresponding AI *versus* explainable artificial intelligence (XAI) outputs.

AI Applications to Medical Imaging

Pros and cons and recommendations for choosing machine learning or deep learning for application to medical imaging.

	Pros	Cons	Recommendations*
ML	• A relatively small sample size can be used • Both discrete and continuous variables for *labelling* are possible, eventually with proper feature *oversampling* • Medical image application domain exists and guides the process • (IBSI standardized features for radiomics) • Integration with additional data is possible and easy • High interpretability is immediately provided by some models (*e.g.*, *decision trees*) and is achievable by other algorithms (*e.g.*, SVM)	• Data curation is particularly time-consuming for *image segmentation* • The model must be selected among the possible algorithms (SVM, *random forest*, Bayesian, etc.)	• *Nested* or *wrapped validation* should be performed • Avoiding dependency on the data *via* careful radiomic feature robustness and reliability analyses to avoid *overfitting* on the development set • Apply feature *harmonization*, *intensity normalization*, *denoising* • List the selected features and the most important or relevant features for the model for explainability
DL	• Learning curve can be used for stopping sample size • Limited samples can be used but with *transfer learning* or eventually with proper *data augmentation* • Suitable for discrete variables for *labelling* • Medical image application domain exists but does not guide the process • (Use *transfer learning* and domain adaptation to take advantage of pretrained models or labelled instances from similar domains) • *Harmonization*, *intensity normalization*, *denoising* could be avoided if images from variety of datasets are present	• Integration with additional data is possible but very complex • Data curation is particularly time-consuming for *labelling* and *annotations* for image *semantic segmentation* • The ML model must be selected among the possible neural network architectures	• Modify architecture to improve the model performance • Use optimizers in training convergence • Use regularization to improve model generalizability • Provide the saliency map of the activated features for explainability

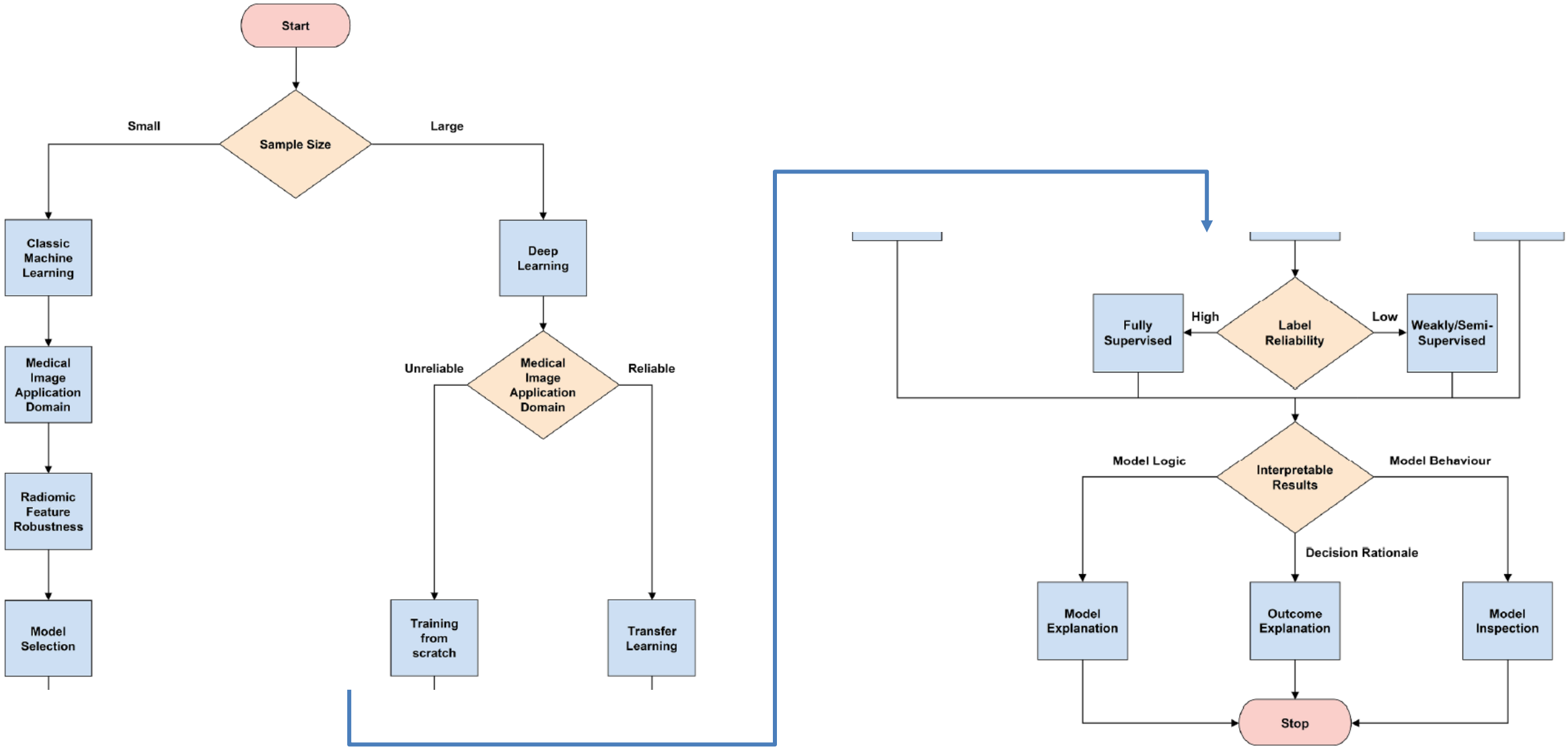
ML = machine learning; DL = deep learning; IBSI = Image Biomarker Standardization Initiative; SVM = support vector machines.
*From a general point of view, ensemble learning can be useful in several situations, and the Vapnik–Chervonenkis method can be help sample size definition.

AI Applications to Medical Imaging

Challenges of classic machine learning and deep learning models according to decision choices.

Challenges	Classic Machine Learning	Deep Learning
Sample size	• Careful radiomic feature robustness and reliability analyses • Strong feature selection process • Machine learning model selection	• *Data augmentation*; *transfer learning* • Regularization to improve model generalizability • *Weakly-*, *semi-*, *self-supervised* or *unsupervised* pre-training • Modify model architecture
Medical image application domain	Avoiding dependency on the data *via* careful radiomic feature robustness analyses to avoid *overfitting* on the development set	Use *transfer learning* and domain adaptation to take advantage of pre-trained models or labelled instances from similar domains
Label and annotation reliability	• Data curation considering both *segmentation* and response variables • To increase the reliability, multiple labels and morphological perturbations could be considered in the feature robustness analyses	• Data curation considering multicentric and multireader study • Use of image-level labels to derive pixel/voxel-level predictions (*inexact supervision*) • Combine a few well-labelled instances with weakly labelled (*inaccurate supervision*) or unlabeled ones (*incomplete supervision*)
Interpretability	High interpretability provided by some models (*e.g.*, *decision trees*) and selected radiomic features (in terms of relevance or importance)	Adopt interpretability and explainability techniques to improve model transparency during both the design and evaluation phases

AI Applications to Medical Imaging



AI Applications to Medical Imaging

Physica Medica 83 (2021) 9–24

Contents lists available at ScienceDirect

Physica Medica

journal homepage: www.elsevier.com/locate/ejmp

Review paper

AI applications to medical images: From machine learning to deep learning

Isabella Castiglioni [a,b,1], Leonardo Rundo [c,d,1], Marina Codari [e,1], Giovanni Di Leo [f], Christian Salvatore [g,h,*], Matteo Interlenghi [h], Francesca Gallivanone [b], Andrea Cozzi [i], Natascha Claudia D'Amico [j,k], Francesco Sardanelli [f,i]

Radiology: Artificial Intelligence

EDITORIAL

Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers

John Mongan, MD, PhD • Linda Moy, MD • Charles E. Kahn, Jr, MD, MS

From the Department of Radiology and Biomedical Imaging, University of California–San Francisco, San Francisco, Calif (J.M.); Department of Radiology and Center for Advanced Imaging Innovation and Research, New York University School of Medicine, New York, NY (L.M.); and Department of Radiology, University of Pennsylvania, 3400 Spruce St, 1 Silverstein, Philadelphia, PA 19104 (C.E.K.). Received March 4, 2020; revision requested March 5; accepted March 5. **Address correspondence to** C.E.K. (e-mail: *ckahn@rsna.org*).

Conflicts of interest are listed at the end of this article.

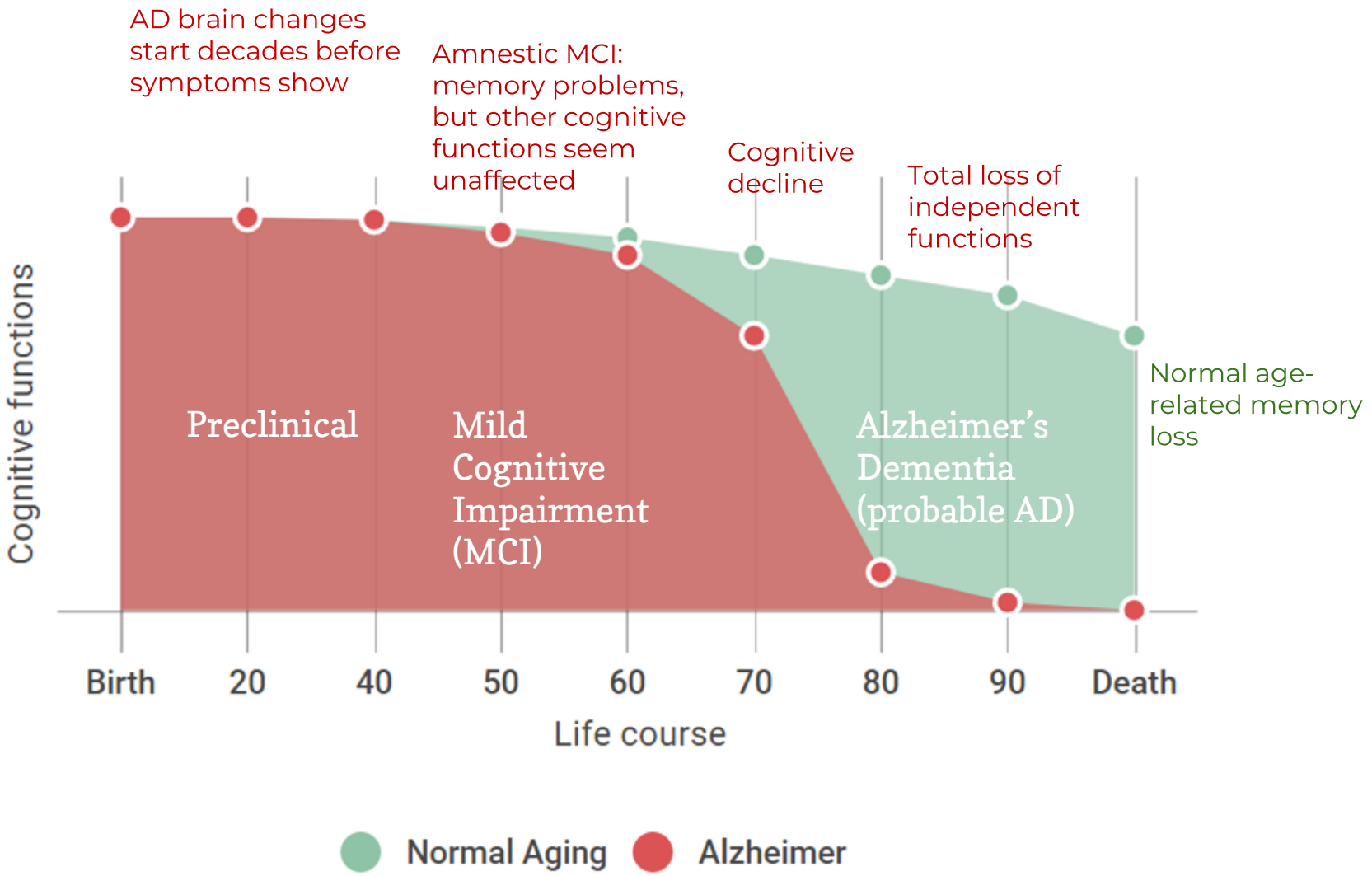
Radiology: Artificial Intelligence 2020; 2(2):e200029 • https://doi.org/10.1148/ryai.2020200029 • Content codes: IN AI • ©RSNA, 2020

ML Models | Examples

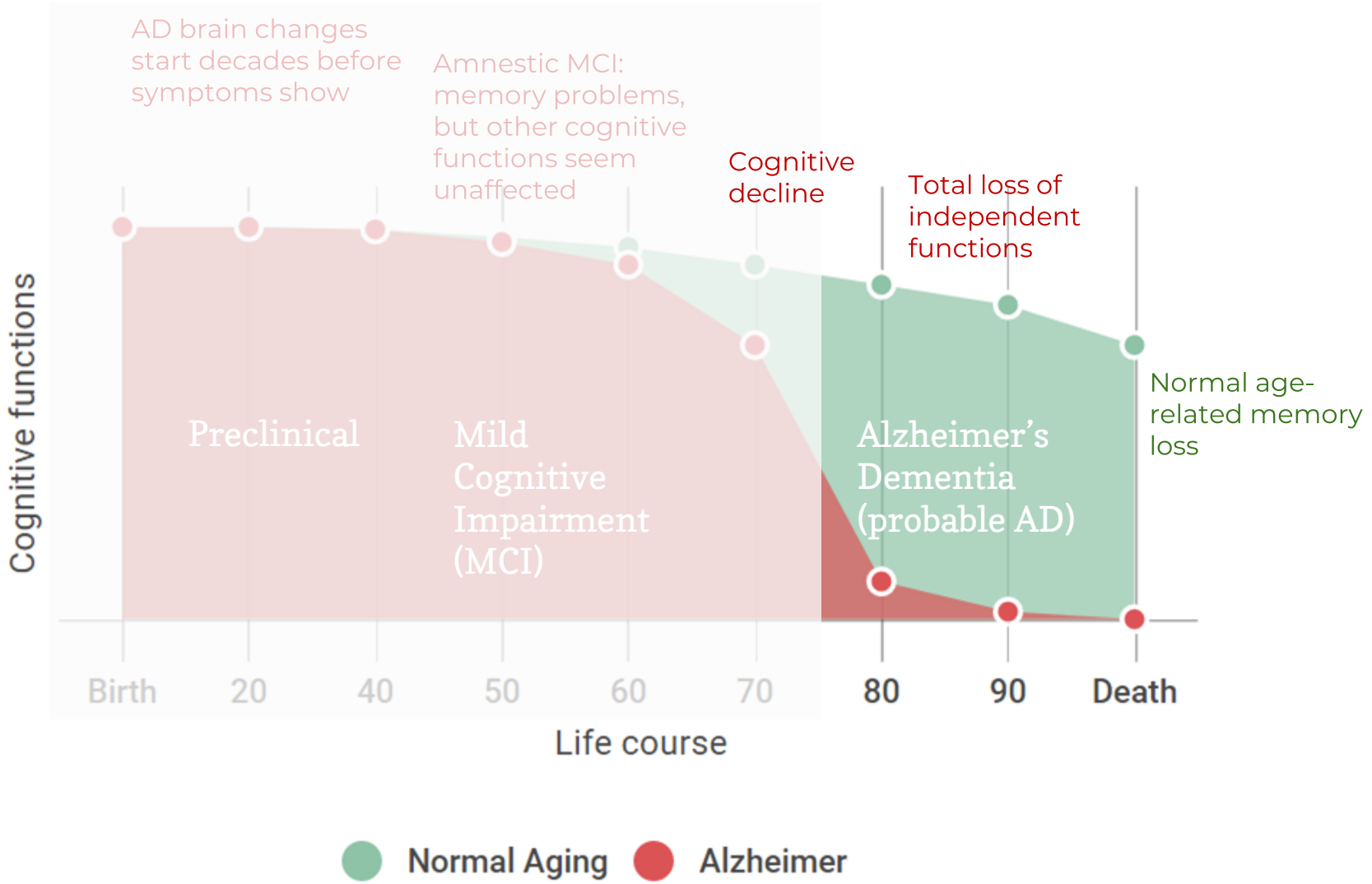
ML models | Examples

Alzheimer's Disease

ML models | Examples

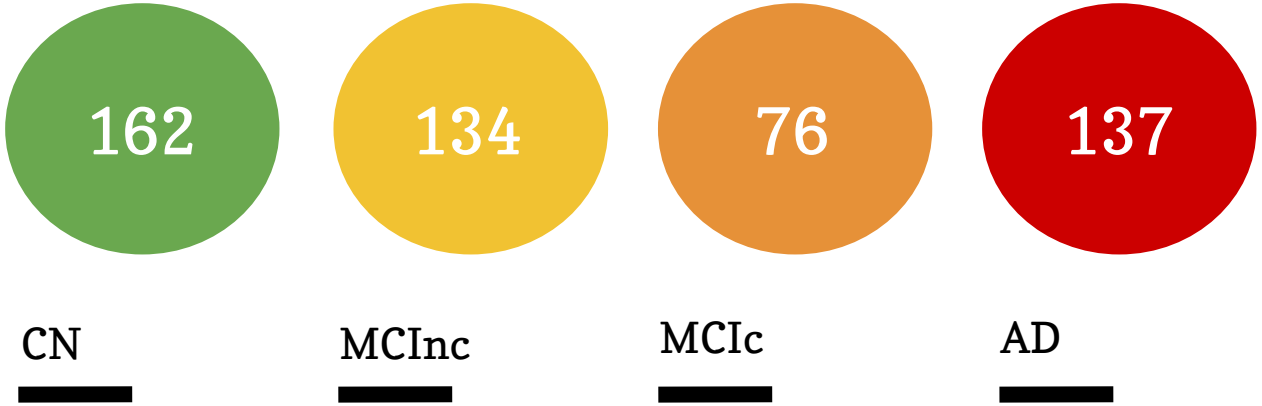


ML models | Examples



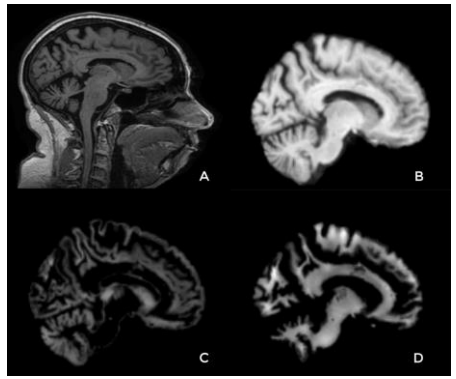
ML models | Examples

Clinical Diagnosis of Alzheimer's Disease



A dataset
of 509 subjects

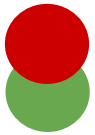
Structural MRI
T1 weighted
1.5 Tesla

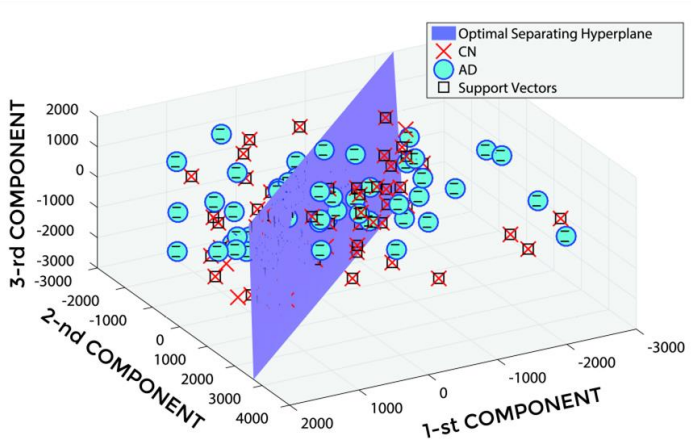


AD	**Alzheimer's Disease**
MCIc	Mild Cognitive Impairment, converting to Alzheimer's Dementia
MCInc	Mild Cognitive Impairment, not converting to Alzheimer's Dementia
CN	Cognitively-Normal subjects

Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: A machine learning approach.
Salvatore et al. 2015, Frontiers in Neuroscience.

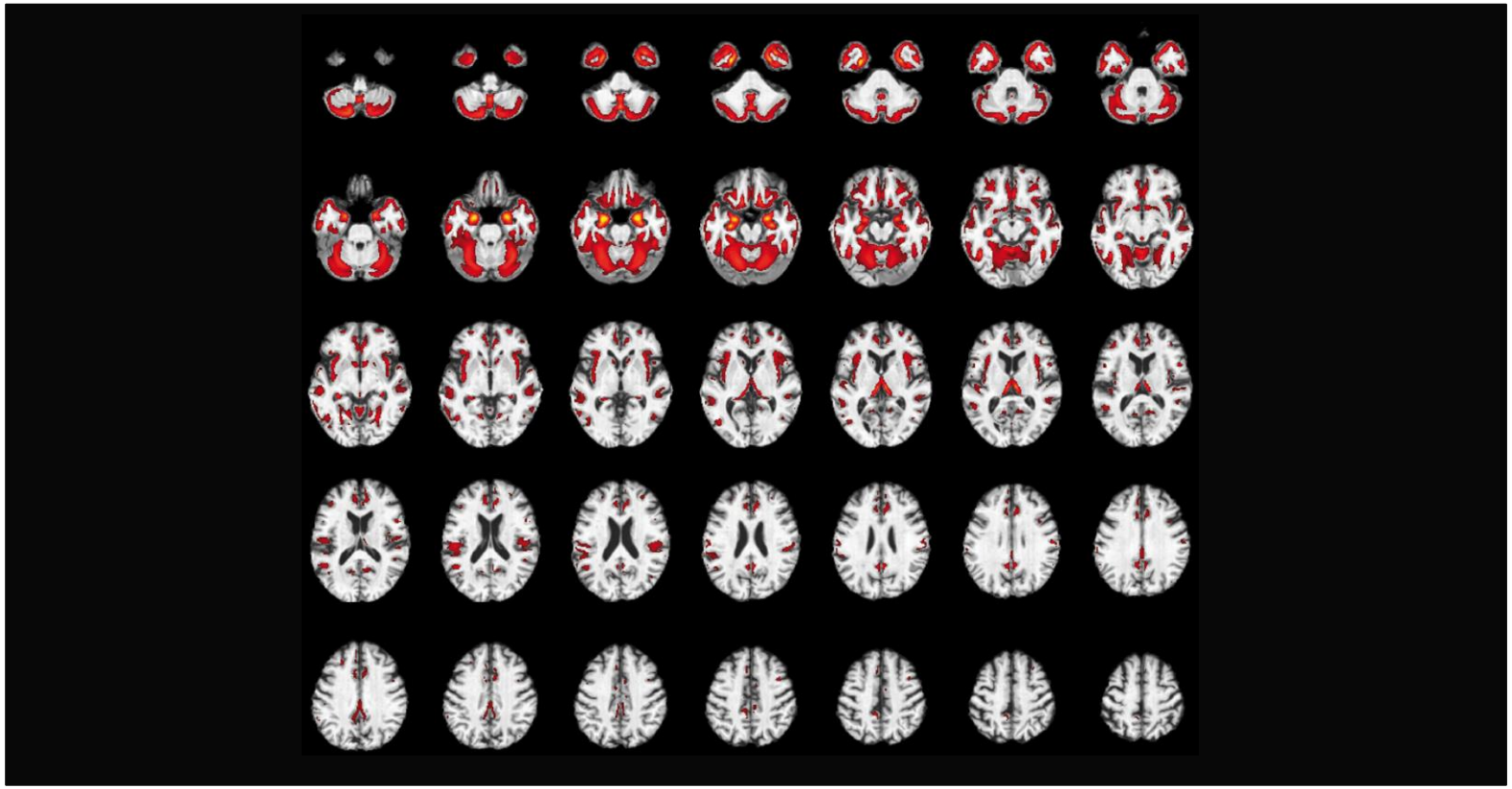
ML models | Examples

	Clinical Diagnosis			Accuracy (automatic classification)
	AD	vs	CN	76%
	MCIc	vs	CN	72%
	MCIc	vs	MCInc	66%



Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: A machine learning approach.
Salvatore et al. 2015, Frontiers in Neuroscience.

ML models | Examples

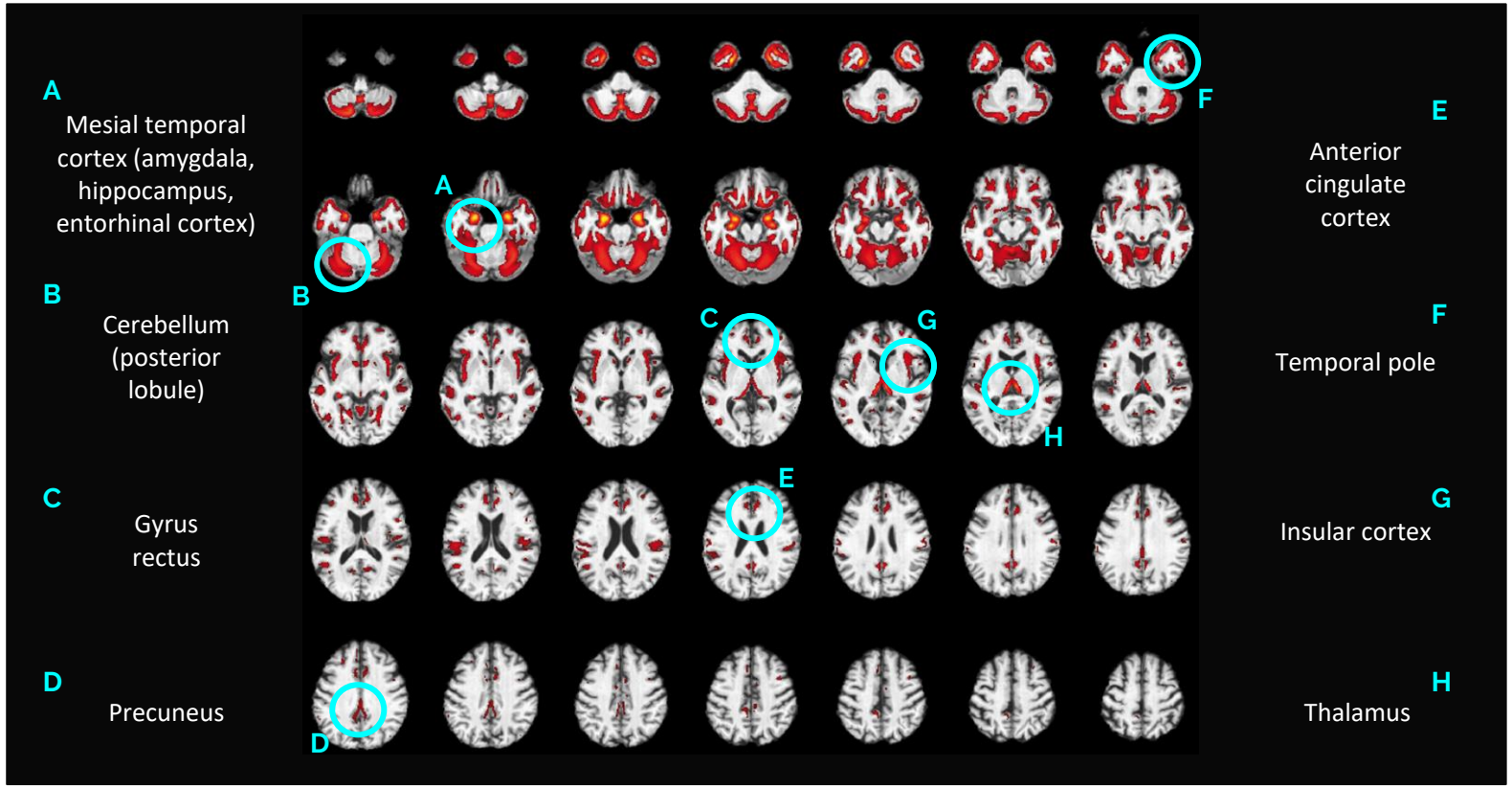


Best Structural-MRI Predictors

AD vs CN

Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: A machine learning approach.
Salvatore et al. 2015, Frontiers in Neuroscience.

ML models | Examples

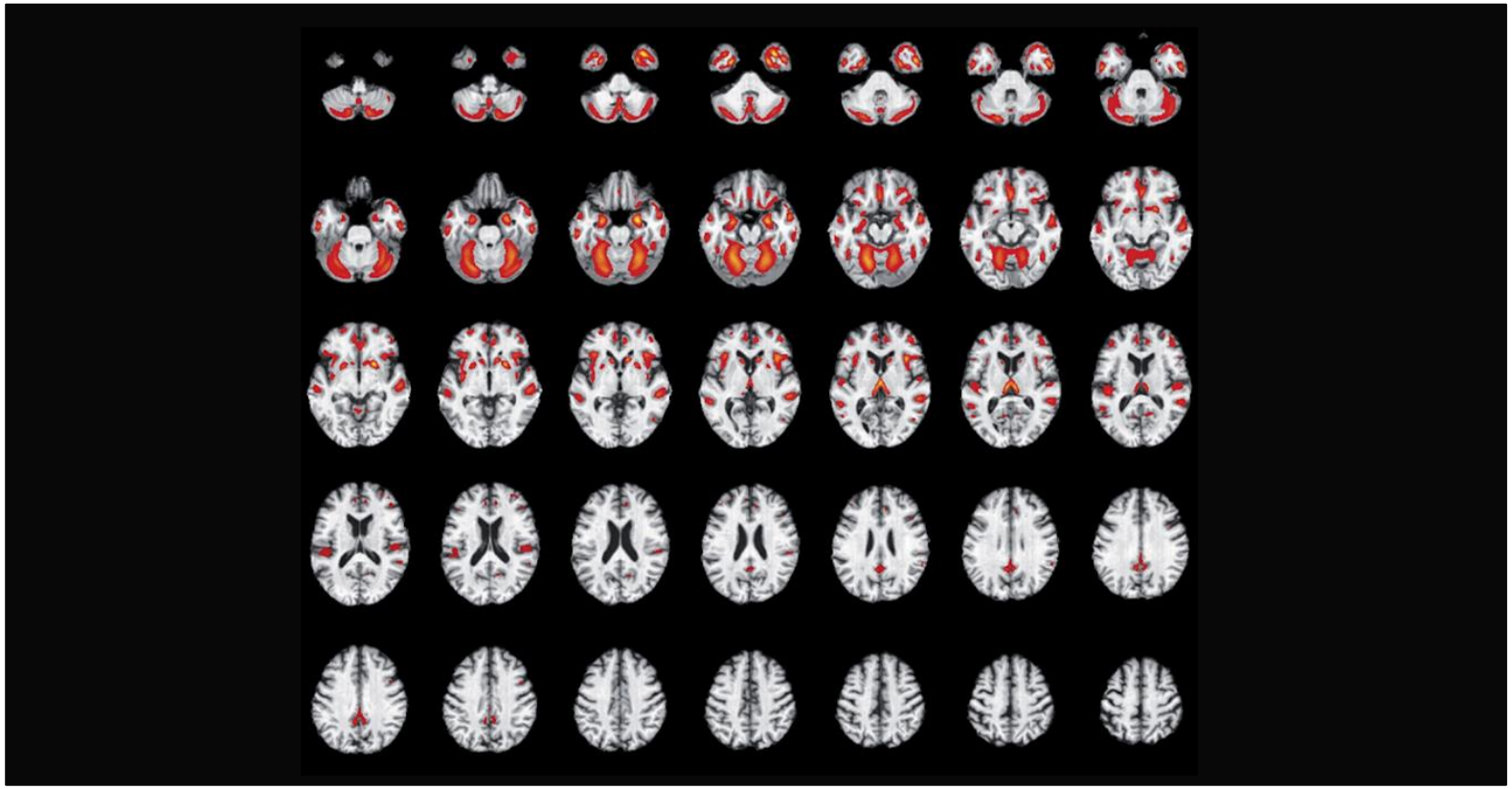


Best Structural-MRI Predictors

AD vs CN

Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: A machine learning approach.
Salvatore et al. 2015, Frontiers in Neuroscience.

ML models | Examples

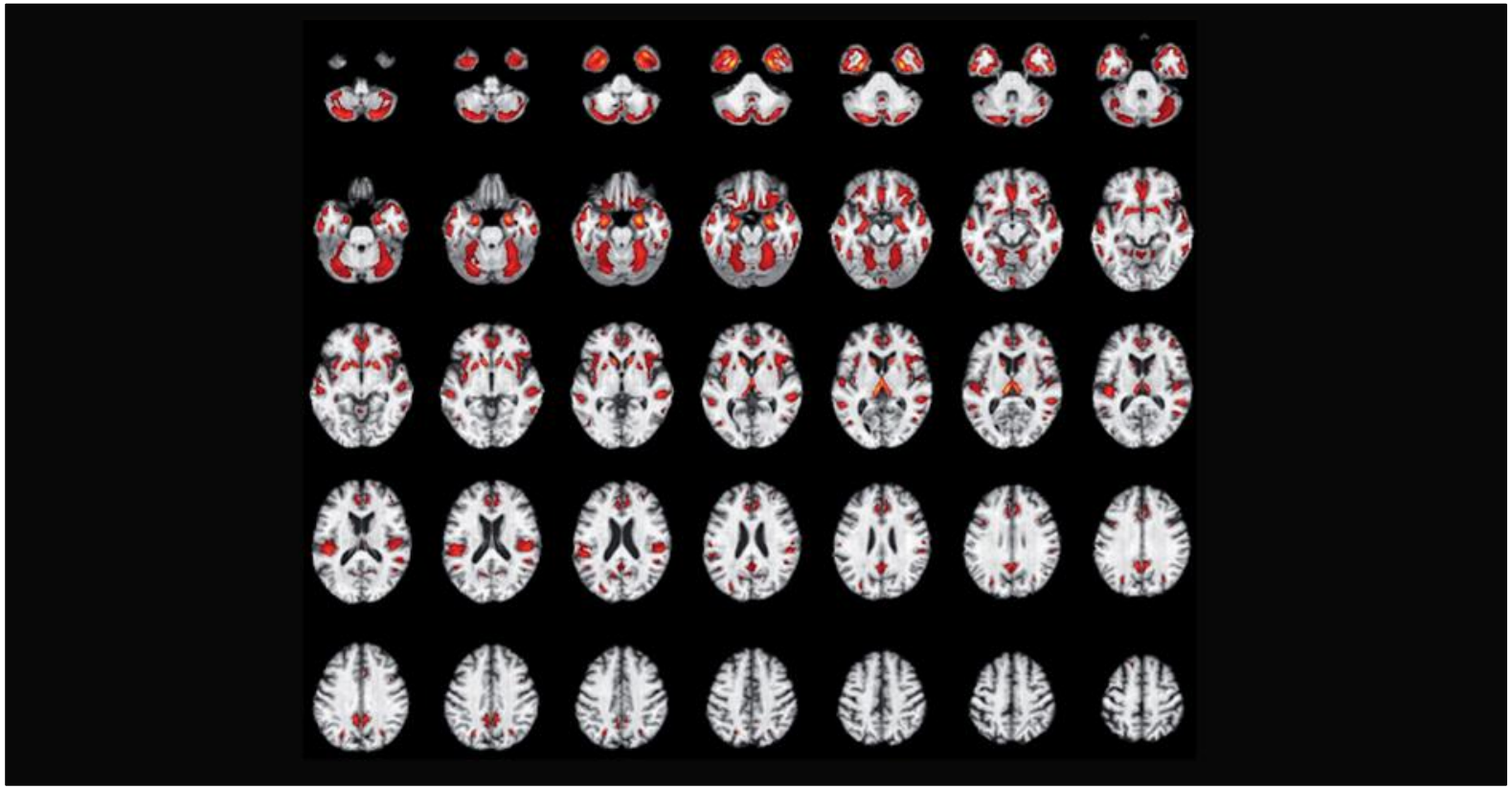


Best Structural-MRI Predictors

MCIc vs CN

Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: A machine learning approach.
Salvatore et al. 2015, Frontiers in Neuroscience.

ML models | Examples



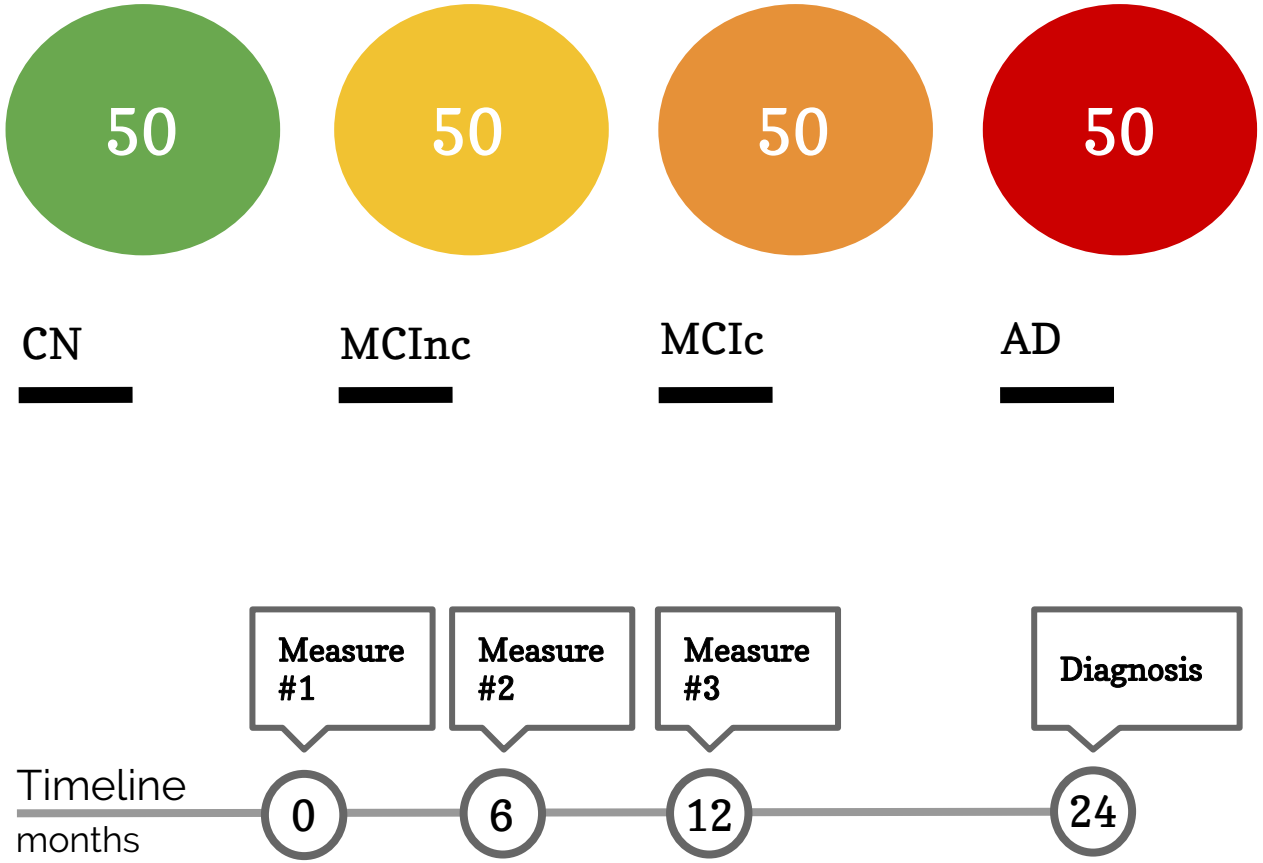
Best Structural-MRI Predictors

MCIc vs MCInc

Magnetic resonance imaging biomarkers for the early diagnosis of Alzheimer's disease: A machine learning approach.
Salvatore et al. 2015, Frontiers in Neuroscience.

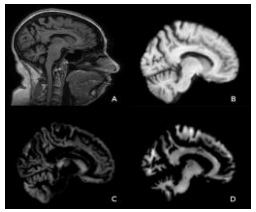
ML models | Examples

Clinical Diagnosis of AD at 24 months follow up



A dataset of 200 subjects

Structural MRI T1 weighted 1.5 Tesla



65 scores/subscores from 7 neuro-psychological tests

MRI characterizes the progressive course of AD and predicts conversion to Alzheimer's dementia twenty-four months before probable diagnosis. *Salvatore et al. 2018, Frontiers in Aging Neuroscience.*

ML models | Examples

Clinical Diagnosis (24 months follow up)

AD + MCIc

vs

MCInc + CN

Accuracy (sen/spe)

79% (79/78)	85% (83/87)
MRI	MRI + NPS

24 months before stable diagnosis

MRI characterizes the progressive course of AD and predicts conversion to Alzheimer's dementia twenty-four months before probable diagnosis. *Salvatore et al. 2018, Frontiers in Aging Neuroscience.*

ML models | Examples

Best Neuropsychological Predictors

24 months before stable diagnosis

Ability in remembering appointments, family occasions, holidays, medications in FAQ
Functional abilities

Ability in writing checks, paying bills, or balancing checkbook in FAQ
Functional abilities

Ability in assembling tax records, business affairs in FAQ
Functional abilities

Total score of trial 5 in AVLT
Memory and learning

Ability in keeping track of current events in FAQ
Functional abilities

Total intrusions of trial 1 in AVLT
Memory and learning

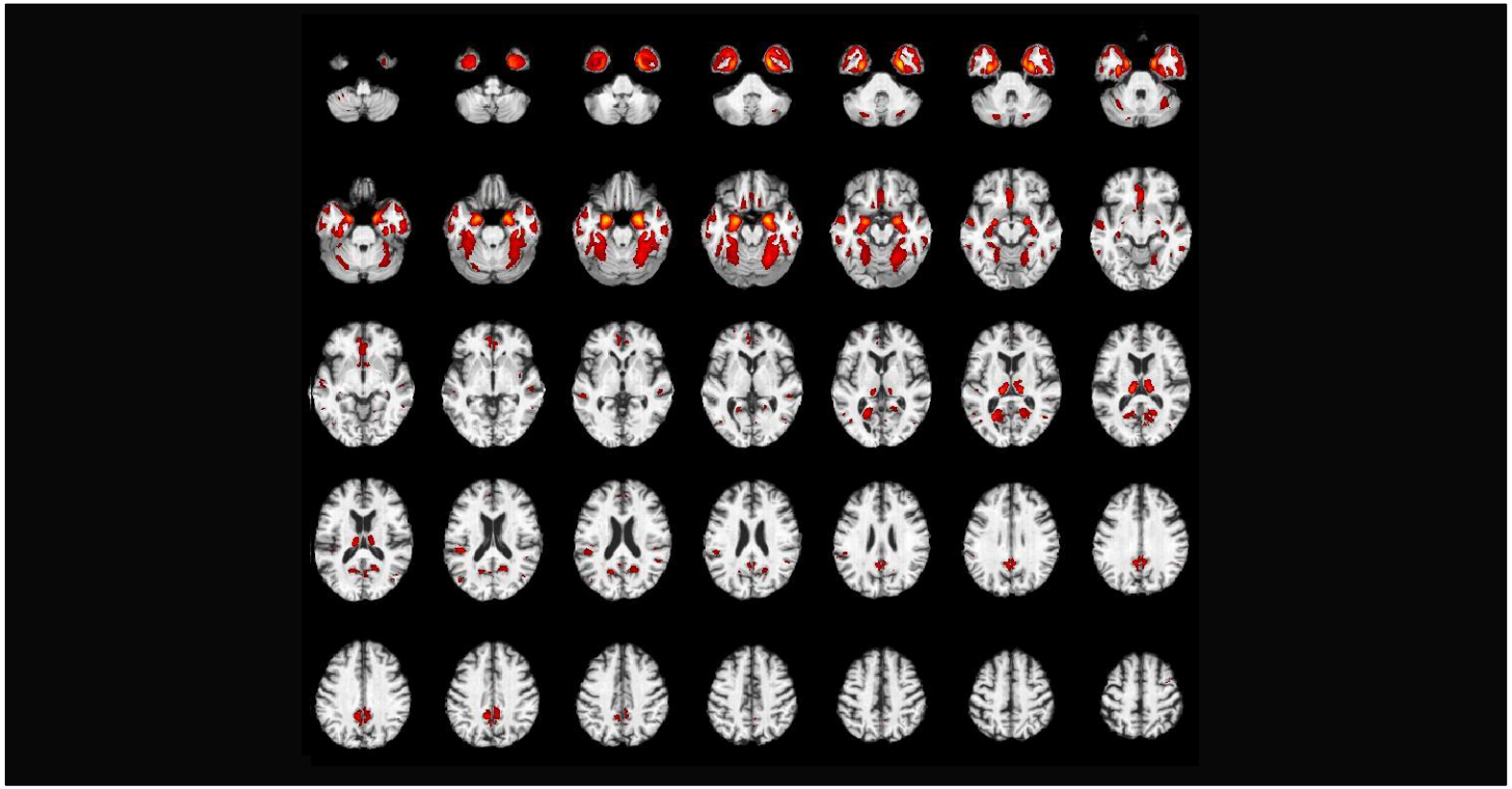
Correct answers in the Backwards task in Digit-Span Test
Working memory

Correct answers in Vegetables task in Category Fluency Test
Language

Correct answers after a 30-min delay in AVLT
Memory and learning

MRI characterizes the progressive course of AD and predicts conversion to Alzheimer's dementia twenty-four months before probable diagnosis. *Salvatore et al. 2018, Frontiers in Aging Neuroscience.*

ML models | Examples

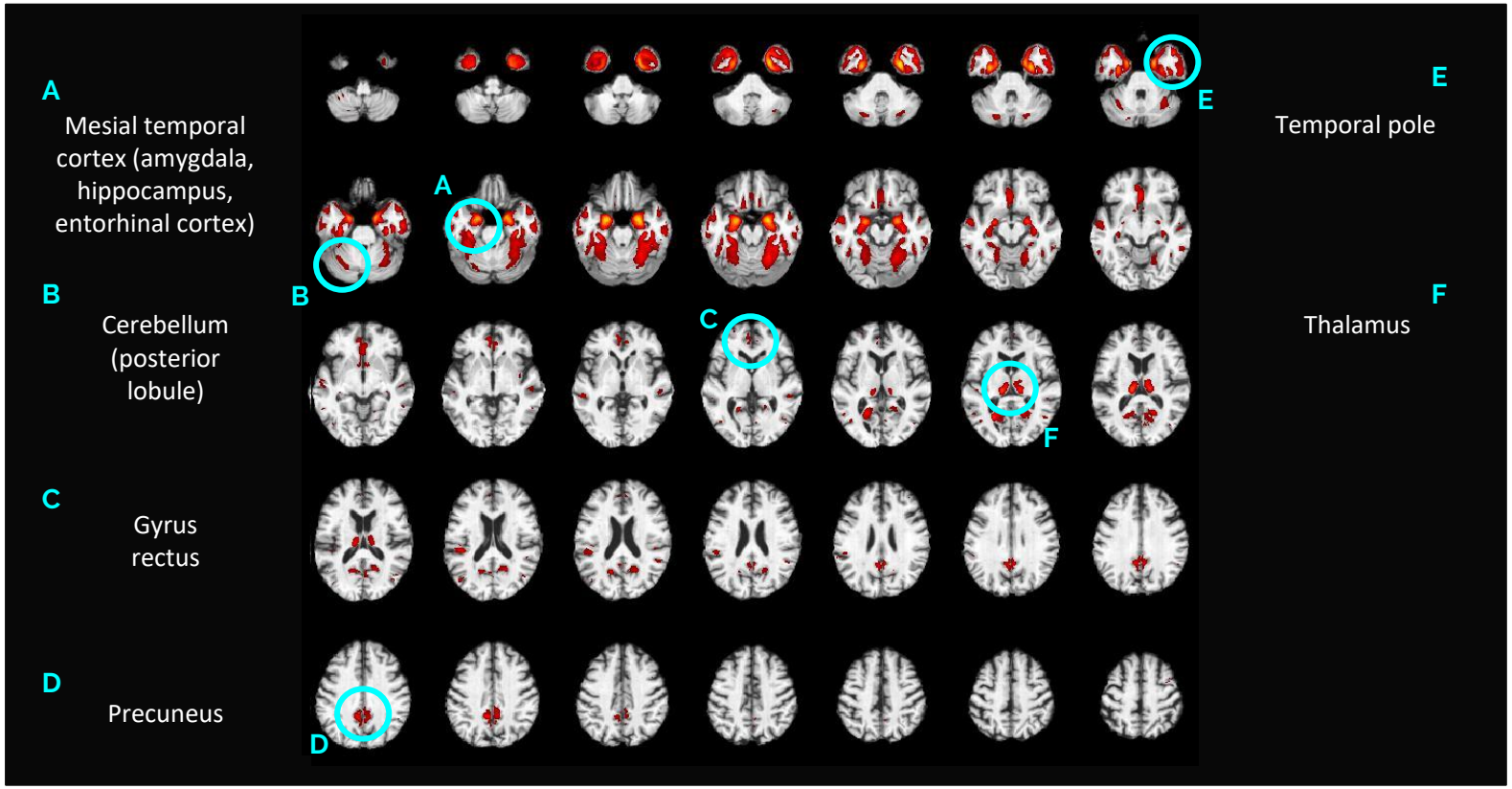


Best Structural-MRI Predictors

24 months before stable diagnosis

MRI characterizes the progressive course of AD and predicts conversion to Alzheimer's dementia twenty-four months before probable diagnosis. *Salvatore et al. 2018, Frontiers in Aging Neuroscience.*

ML models | Examples



Best Structural-MRI Predictors

24 months before stable diagnosis

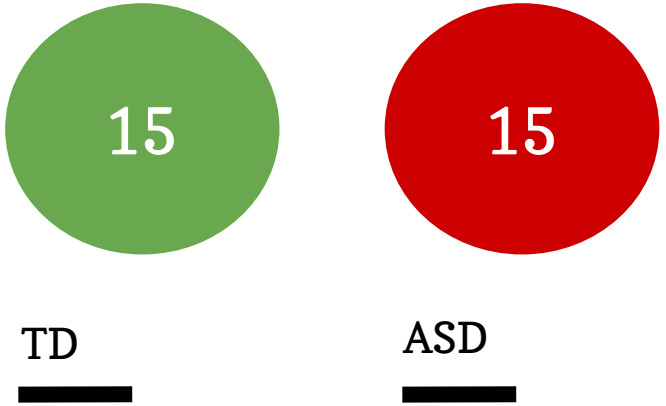
MRI characterizes the progressive course of AD and predicts conversion to Alzheimer's dementia twenty-four months before probable diagnosis. *Salvatore et al. 2018, Frontiers in Aging Neuroscience.*

ML models | Examples

Autism Spectrum Disorder

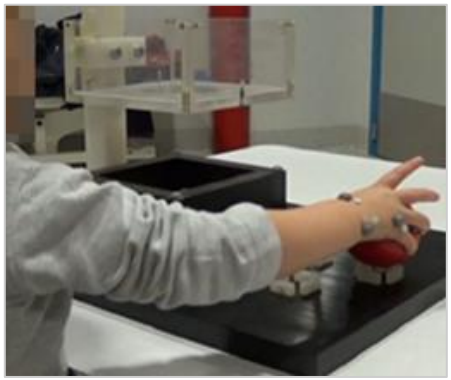
ML models | Examples

Confirm a motor signature of autism



A dataset of 30
pre-school children (~3 years old)

17 kinematic features
collected during a reach-to-drop task



ASD **Autism Spectrum Disorder**
TD Typically-developing children

Use of machine learning to identify children with autism and their motor abnormalities. *Crippa, Salvatore et al. 2015, Journal of Autism and Developmental Disorders.*

ML models | Examples

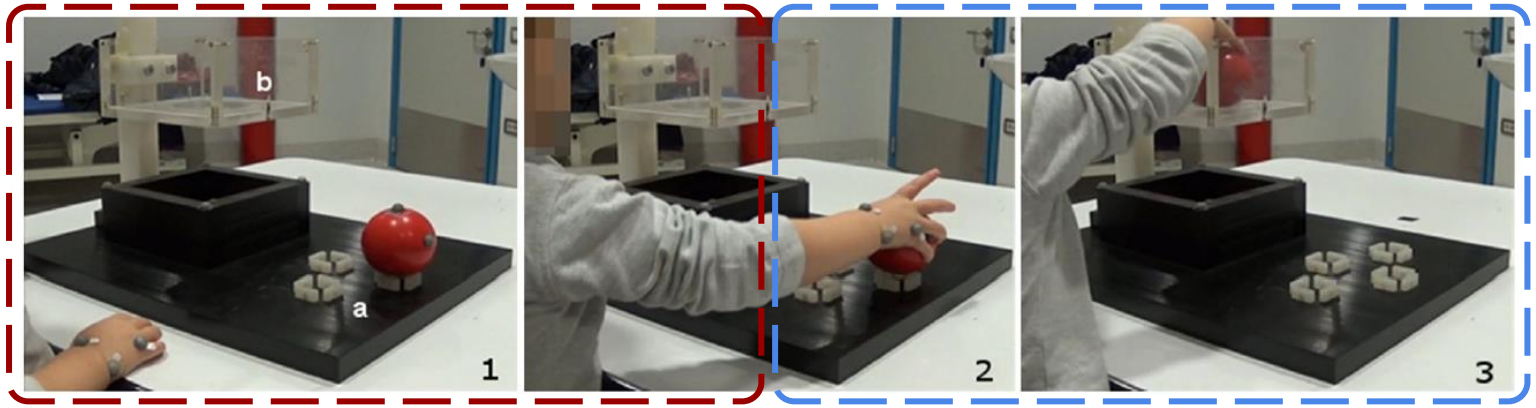


Fig. 1 The experimental task consisted of grasping a rubber ball (**2**) that was placed over a support (see **1**, a); that is, a reach-to-grasp movement before they dropped it in a hole (**3**). The hole (**1**, c) was located inside a see-through *square box* (21 cm high, 20 cm wide) and was large enough not to require fine movements. The goal area is transparent to allow seeing through. 4 markers are placed on the basket under the goal area, 2 on the ball and 3 on each hand (attached to the ulnar and radial surfaces of the participant's wrist and to the hand dorsum on the 4th and 5th metacarpals)

sub-movement 1
the movement necessary to reach the ball and place it on its support

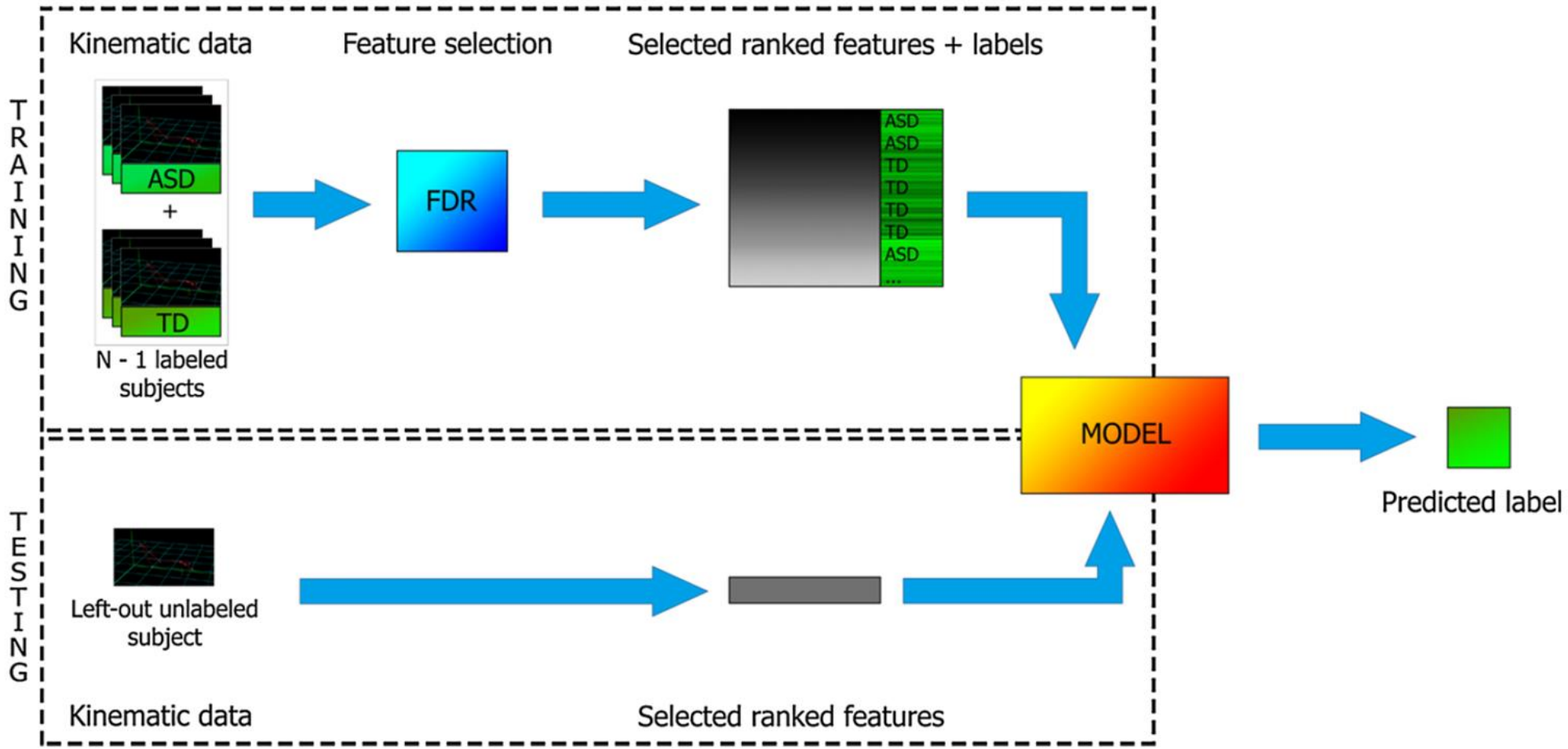
- # movement units
- total movement duration
- peak velocity
- peak acceleration
- time of peak acceleration
- peak deceleration
- time of peak deceleration

sub-movement 2
the movement to transport the ball from a support to the target hole

- # movement units
- total movement duration
- peak velocity
- peak acceleration
- time of peak acceleration
- peak deceleration
- time of peak deceleration
- wrist angle

Use of machine learning to identify children with autism and their motor abnormalities. *Crippa, Salvatore et al. 2015, Journal of Autism and Developmental Disorders.*

ML models | Examples



Use of machine learning to identify children with autism and their motor abnormalities. *Crippa, Salvatore et al. 2015, Journal of Autism and Developmental Disorders.*

ML models | Examples

	Overall mean	Optimal configuration
Diagnostic accuracy (sensitivity / specificity)	85 (82/89)%	97 (100/94)%

Use of machine learning to identify children with autism and their motor abnormalities. *Crippa, Salvatore et al. 2015, Journal of Autism and Developmental Disorders.*

ML models | Examples

	Overall mean	Optimal configuration
Diagnostic accuracy (sensitivity / specificity)	85 (82/89)%	97 (100/94)%

7 optimal features out of 17

sub-movement 2
the movement to transport the ball from a support to the target hole in which the ball was to be dropped

1. **total movement duration**
2. **delta wrist angle**
3. # movement units
4. time of peak deceleration
5. peak acceleration
6. time of peak velocity
7. peak velocity

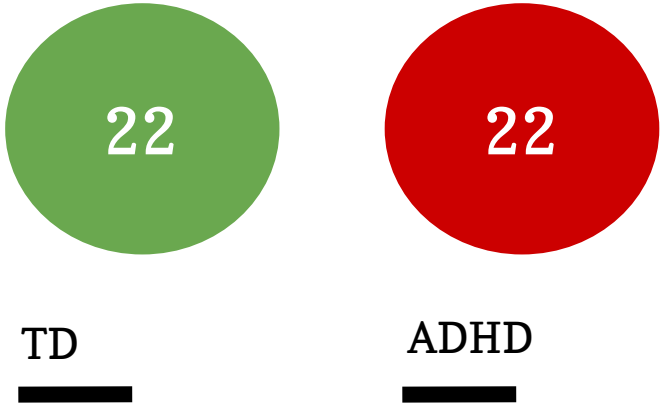
Use of machine learning to identify children with autism and their motor abnormalities. *Crippa, Salvatore et al. 2015, Journal of Autism and Developmental Disorders.*

ML models | Examples

Attention-Deficit/Hyperactivity Disorder

ML models | Examples

Diagnosis of ADHD & Identification of a signature



A dataset of 44
school-aged children (~11 years old)

Multidomain profile of measures
. biological data
. neuropsychological data
. Near-InfraRed Spectroscopy (NIRS)

ADHD **Attention Deficit / Hyperactivity Disorder**
TD Typically-developing children

The Utility of a computerized algorithm Based on a Multi-Domain Profile of Measures for the Diagnosis of attention deficit/hyperactivity Disorder. *Crippa, Salvatore et al. 2017, Frontiers in Psychiatry.*

ML models | Examples

Multi-domain profile of measures

Biological data

10 features

Neuropsychological data

18 features

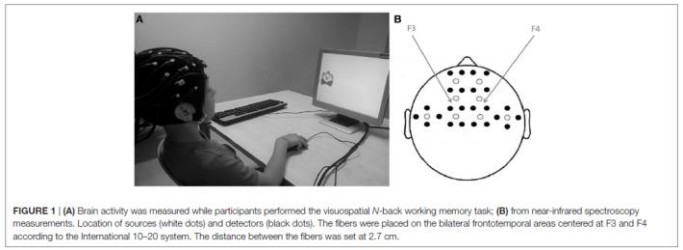


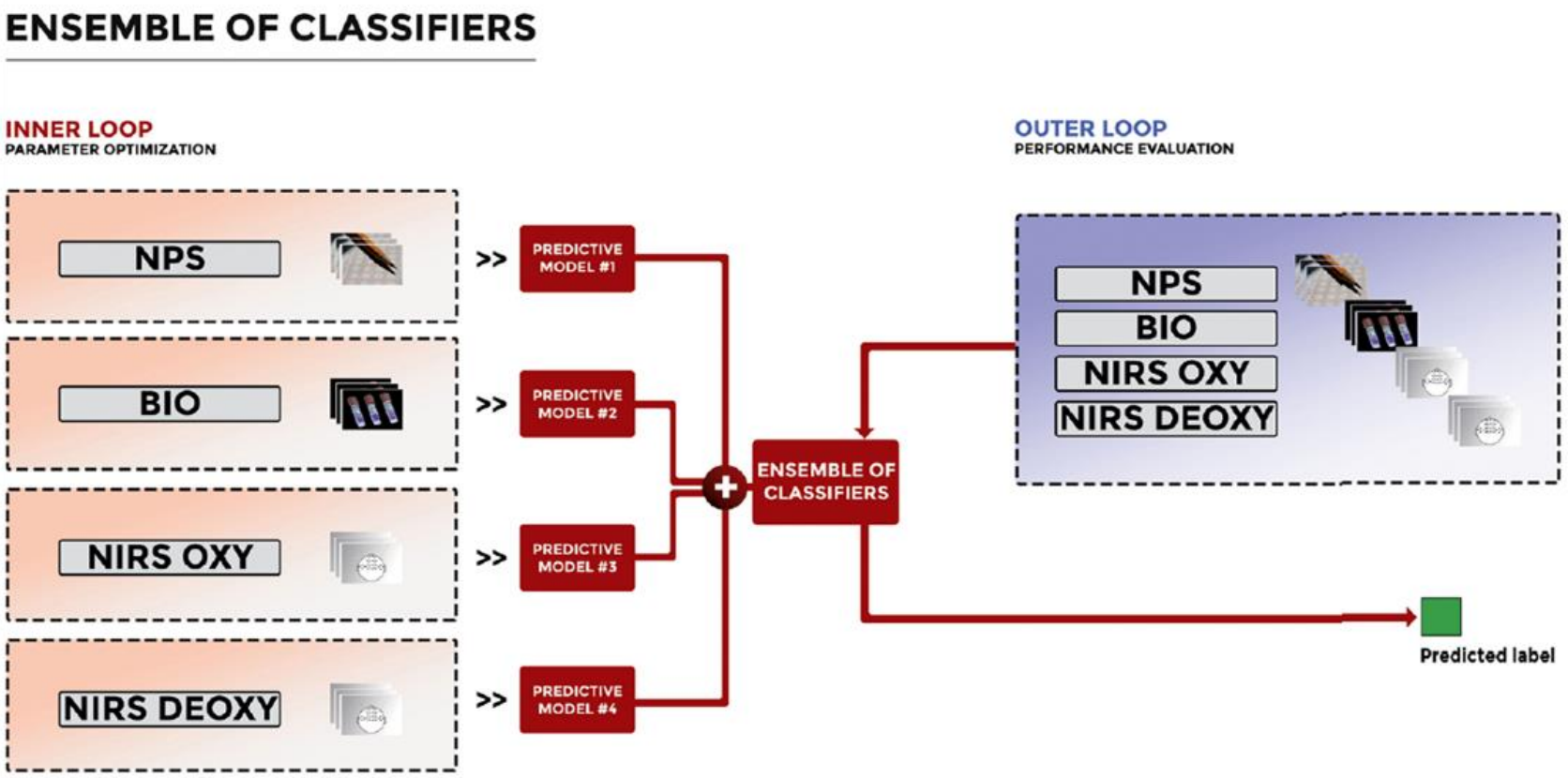
FIGURE 1 | (A) Brain activity was measured while participants performed the visuospatial *N*-back working memory task; (B) from near-infrared spectroscopy measurements. Location of sources (white dots) and detectors (black dots). The fibers were placed on the bilateral frontotemporal areas centered at F3 and F4 according to the International 10–20 system. The distance between the fibers was set at 2.7 cm.

Near-InfraRed Spectroscopy (NIRS)

Oxy/Deoxy data from 32 channels

The Utility of a computerized algorithm Based on a Multi-Domain Profile of Measures for the Diagnosis of attention deficit/hyperactivity Disorder. *Crippa, Salvatore et al. 2017, Frontiers in Psychiatry.*

ML models | Examples



The Utility of a computerized algorithm Based on a Multi-Domain Profile of Measures for the Diagnosis of attention deficit/hyperactivity Disorder. *Crippa, Salvatore et al. 2017, Frontiers in Psychiatry.*

ML models | Examples

Measures	Accuracy (mean ± sd)	Sensitivity (mean ± sd)	Specificity (mean ± sd)
Neuropsychological	62 ± 17	70 ± 27	57 ± 24
Biological	66 ± 21	58 ± 40	73 ± 29
NIRS OXY	57 ± 27	48 ± 47	67 ± 33
NIRS DEOXY	78 ± 22	72 ± 34	82 ± 24
NIRS OXY + DEOXY	72 ± 32	73 ± 29	68 ± 43

biological features

.linoleic acid
.PUFA
.AA
.EPA
.omega-3 index
.AA/DHA
.AA/EPA
.MUFA

linoleic acid and total amount of polyunsaturated fatty acids

The Utility of a computerized algorithm Based on a Multi-Domain Profile of Measures for the Diagnosis of attention deficit/hyperactivity Disorder. *Crippa, Salvatore et al. 2017, Frontiers in Psychiatry.*

ML models | Examples

Measures	Accuracy (mean ± sd)	Sensitivity (mean ± sd)	Specificity (mean ± sd)
Neuropsychological	62 ± 17	70 ± 27	57 ± 24
Biological	66 ± 21	58 ± 40	73 ± 29
NIRS OXY	57 ± 27	48 ± 47	67 ± 33
NIRS DEOXY	78 ± 22	72 ± 34	82 ± 24
NIRS OXY + DEOXY	72 ± 32	73 ± 29	68 ± 43

neuropsychological features

- ***sustained attention***-*false alarms*
- ***visual set***-*shifting-RT inhibition*
- *sustained attention-coefficient of variation*
- *visual set-shifting-number of inhibition errors*
- ***focused attention***-*RT correct responses*
- *focused attention-correct rejections target non-relevant position*
- *focused attention-SD of correct responses RT*
- *focused attention-misses*
- ...

measures of vigilance, focused and sustained attention, and cognitive flexibility

The Utility of a computerized algorithm Based on a Multi-Domain Profile of Measures for the Diagnosis of attention deficit/hyperactivity Disorder. *Crippa, Salvatore et al. 2017, Frontiers in Psychiatry.*

ML models | Examples

Single domain

Measures	Accuracy (mean ± sd)	Sensitivity (mean ± sd)	Specificity (mean ± sd)
Neuropsychological	62 ± 17	70 ± 27	57 ± 24
Biological	66 ± 21	58 ± 40	73 ± 29
NIRS OXY	57 ± 27	48 ± 47	67 ± 33
NIRS DEOXY	78 ± 22	72 ± 34	82 ± 24
NIRS OXY + DEOXY	72 ± 32	73 ± 29	68 ± 43

Multi-domain

Measures	Accuracy (mean ± sd)	Sensitivity (mean ± sd)	Specificity (mean ± sd)
NPS + BIO + NIRS OXY			
NPS + BIO + NIRS DEOXY			
NPS + NIRS OXY + NIRS DEOXY			
BIO + NIRS OXY + NIRS DEOXY			
NPS + BIO + NIRS OXY + NIRS DEOXY			

The Utility of a computerized algorithm Based on a Multi-Domain Profile of Measures for the Diagnosis of attention deficit/hyperactivity Disorder. *Crippa, Salvatore et al. 2017, Frontiers in Psychiatry.*

ML models | Examples

Single domain

Measures	Accuracy (mean ± sd)	Sensitivity (mean ± sd)	Specificity (mean ± sd)
Neuropsychological	62 ± 17	70 ± 27	57 ± 24
Biological	66 ± 21	58 ± 40	73 ± 29
NIRS OXY	57 ± 27	48 ± 47	67 ± 33
NIRS DEOXY	78 ± 22	72 ± 34	82 ± 24
NIRS OXY + DEOXY	72 ± 32	73 ± 29	68 ± 43

Multi-domain

Measures	Accuracy (mean ± sd)	Sensitivity (mean ± sd)	Specificity (mean ± sd)
NPS + BIO + NIRS OXY	71 ± 10	70 ± 27	73 ± 24
NPS + BIO + NIRS DEOXY	81 ± 15	73 ± 24	87 ± 22
NPS + NIRS OXY + NIRS DEOXY	78 ± 18	70 ± 36	87 ± 22
BIO + NIRS OXY + NIRS DEOXY	77 ± 21	63 ± 31	90 ± 21
NPS + BIO + NIRS OXY + NIRS DEOXY	76 ± 16	83 ± 22	68 ± 23

The Utility of a computerized algorithm Based on a Multi-Domain Profile of Measures for the Diagnosis of attention deficit/hyperactivity Disorder. *Crippa, Salvatore et al. 2017, Frontiers in Psychiatry.*

christian.salvatore@iusspavia.it

https://christiansalvatore.github.io/machinelearning-iuss/