



HAL
open science

Design of data driven decision support systems for the early detection of subjects at risk to develop Alzheimer's disease

Manon Ansart

► **To cite this version:**

Manon Ansart. Design of data driven decision support systems for the early detection of subjects at risk to develop Alzheimer's disease. Machine Learning [stat.ML]. Sorbonne Université, 2019. English. NNT: . tel-02425735

HAL Id: tel-02425735

<https://theses.hal.science/tel-02425735>

Submitted on 31 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sorbonne Université

École Doctorale d'Informatique, de Télécommunication et d'Électronique (EDITE)

ARAMIS LAB à l'Institut du Cerveau et de la Moelle épinière (ICM)

DESIGN OF DATA DRIVEN DECISION SUPPORT SYSTEMS FOR THE EARLY DETECTION OF SUBJECTS AT RISK TO DEVELOP ALZHEIMER'S DISEASE

CREATION DE SYSTEMES D'AIDE A LA DECISION POUR LA DETECTION PRECOCE DE SUJETS A RISQUE
DE DEVELOPPER LA MALADIE D'ALZHEIMER

MANON ANSART

Thèse de doctorat d'informatique

Dirigée par *Didier Dormont* et *Stanley Durrleman*

Présentée et soutenue publiquement le 25 novembre 2019

Devant un jury composé de :

- *Emmanuel Bacry*
Directeur de Recherche, CNRS, Examineur
- *Didier Dormont*
Professeur, APHP, Directeur de thèse
- *Stanley Durrleman*
Directeur de Recherche, Inria, Directeur de thèse
- *Julie Josse*
Professeur, École Polytechnique, Examinatrice
- *Christophe Marsala*
Professeur, Sorbonne Université, Examineur
- *Pierre Pudlo*
Professeur, Aix-Marseille Université, Rapporteur
- *Rodolphe Thiébaud*
Directeur de Recherche, Inserm, Rapporteur

Abstract

DESIGN OF DATA DRIVEN DECISION SUPPORT SYSTEMS FOR THE EARLY DETECTION OF SUBJECTS AT RISK TO DEVELOP ALZHEIMER'S DISEASE

by Manon ANSART

The goal of this thesis is to design data-driven methods to identify subjects at risk to develop Alzheimer's disease. As it is a progressive disease, subtle signs can appear several years before the first clinical symptoms. Identifying subjects who show these signs, and who are likely to develop the disease in the coming years, is a crucial point that could allow researchers to better study the disease mechanism, select patients for clinical trials and tailor patient care.

In the first chapter, we conduct a review of methods predicting the future diagnosis of subjects suffering from mild cognitive impairment. We quantitatively and qualitatively study these methods, and take a critical view point by identifying several methodological issues. In the second chapter, we propose our own method to predict the future diagnosis by using a two-step approach : we first predict the future subject characteristics, and then use this result to predict the corresponding diagnosis. In the third chapter, we propose an automatic method to select subjects with a positive biomarker for clinical trials, so as to minimize the recruitment cost. In the last chapter, we analyze prescription patterns before and after diagnosis using a medical record database. We use them to predict if a patient will develop Alzheimer's disease in the next five or ten years.

Across these works, we show the importance to take into account the adoption of these methods and the settings in which they can be used, especially regarding the test cohort, the data types and the interpretability of the method.

Résumé

CREATION DE SYSTEMES D'AIDE A LA DECISION POUR LA DETECTION PRECOCE DE SUJETS A RISQUE DE DEVELOPPER LA MALADIE D'ALZHEIMER

by Manon ANSART

Le but de cette thèse est de proposer des méthodes d'apprentissage automatique pour identifier des sujets à risque de développer la maladie d'Alzheimer. L'identification à un stade très précoce de sujets à risque de développer la maladie est une problématique clé, qui permettrait de mieux étudier la maladie, de sélectionner des patients pour des essais cliniques et de leur proposer un suivi adapté.

Dans un premier chapitre, nous effectuons une revue des méthodes prédisant le diagnostic futur de sujets atteints de troubles cognitifs légers. Nous effectuons un travail de synthèse, à la fois qualitatif et quantitatif, des méthodes proposées pour effectuer cette prédiction et des problèmes méthodologiques qu'elles comportent. Dans un deuxième chapitre, nous proposons d'effectuer cette prédiction du futur diagnostic avec une approche en deux temps : nous prédisons d'abord l'évolution des caractéristiques des sujets, et utilisons ces résultats pour prédire le diagnostic correspondant à un stade ultérieur. Dans un troisième chapitre, nous proposons une méthode automatique permettant de repérer des sujets à biomarqueurs positifs pour les essais cliniques, de manière à minimiser le coût de recrutement. Dans un dernier chapitre, nous analysons l'évolution des prescriptions de médicaments avant et après le diagnostic grâce à des bases d'historiques médicaux. Nous les utilisons pour prédire si un patient va développer la maladie d'Alzheimer dans les 5 ou 10 années à venir.

Nous mettons en avant l'importance de prendre en compte l'adoption des méthodes et leur cadre d'utilisation, notamment à travers la cohorte d'étude, les types de données, et l'interprétabilité de la méthode.

Remerciements

J'aimerais tout d'abord remercier mes directeurs de thèse, Prof. Didier Dormont et Stanley Durrleman, pour m'avoir donné la possibilité de faire ce doctorat. Je remercie tout particulièrement Stanley pour m'avoir guidée pendant ces années, tout en me laissant la liberté d'explorer les sujets qui m'intéressaient le plus, sous un angle qui m'était propre.

Je remercie l'Inserm et le projet EuroPOND pour avoir financé ces travaux, et Cegedim pour m'avoir donné accès à leur base de données, me permettant de donner une autre dimension à ces recherches. Je remercie Laurene Gantzer et Béranger Lekens, pour m'avoir conseillée et guidée dans l'utilisation de ces données.

Je suis très reconnaissante envers les chercheurs travaillant à Aramis, pour m'avoir apporté leur aide au cours de ces années. Je remercie en particulier Stéphane Epelbaum, pour avoir apporté en tant que neurologue une expertise médicale et une connaissance de la pratique clinique qui ont donné tout leur intérêt à ces travaux; Olivier Colliot, qui m'a conseillée dans la réalisation et la rédaction de plusieurs études; Ninon Burgos, pour m'avoir offert son soutien et guidée dans les méandres de la thèse à plusieurs reprises.

Je remercie également toutes les personnes m'ayant accompagnée pendant ces années de thèse. Je remercie tout particulièrement mes collègues de l'équipe Aramis, qui ont participé à rendre ces années agréables, ainsi que les doctorants qui ont traversé avec moi ces années riches en expériences, en particulier Tilia Dahou.

Pour finir, j'aimerais remercier ma famille pour m'avoir encouragée et soutenue, et Antoine, pour son amour et son soutien indéfectible.

Scientific production

JOURNAL PAPERS

1. **Manon Ansart**, Stéphane Epelbaum, Geoffroy Gagliardi, Olivier Colliot, Didier Dormont, Bruno Dubois, Harald Hampel, Stanley Durrleman, and for the Alzheimer's Disease Neuroimaging Initiative and the INSIGHT-preAD study. 2019. "Reduction of Recruitment Costs in Preclinical AD Trials : Validation of Automatic Pre-Screening Algorithm for Brain Amyloidosis." **Statistical Methods in Medical Research**, January, 096228021882303. <https://doi.org/10.1177/0962280218823036>.

SUBMITTED JOURNAL PAPERS

1. **Manon Ansart**, Stéphane Epelbaum, Giulia Bassignana, Alexandre Bône, Simona Botani, Tiziana Cattai, Raphaël Couronné, Johann Faouzi, Igor Koval, Maxime Louis, Elina Thibeau-Sutre, Junhao Wen, Adam Wild, Ninon Burgos, Didier Dormont, Olivier Colliot, Stanley Durrleman. "Predicting the Progression of Mild Cognitive Impairment Using Machine Learning : A Systematic and Quantitative Review." Submitted in **Medical Image Analysis**.
2. **Manon Ansart**, Stéphane Epelbaum, Béranger Lekens, Laurène Gantzer, Didier Dormont, Stanley Durrleman. "Use of psychotropic drugs throughout the course of Alzheimer's disease : a large-scale study of French medical records" Submitted in **JAMA Neurology**.

PEER-REVIEWED CONFERENCE PROCEEDINGS ---

1. **Manon Ansart**, Stéphane Epelbaum, Geoffroy Gagliardi, Olivier Colliot, Didier Dormont, Bruno Dubois, Harald Hampel, and Stanley Durrleman. 2017. "Prediction of Amyloidosis from Neuropsychological and MRI Data for Cost Effective Inclusion of Pre-Symptomatic Subjects in Clinical Trials." In **Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support**, edited by M. Jorge Cardoso, Tal Arbel, Gustavo Carneiro, Tanveer Syeda-Mahmood, João Manuel R.S. Tavares, Mehdi Moradi, Andrew Bradley, et al., 10553 :357–64. Cham : Springer International Publishing. https://doi.org/10.1007/978-3-319-67558-9_41.

CONFERENCE ABSTRACTS ---

1. **Manon Ansart**, Ninon Burgos, Olivier Colliot, Didier Dormont and Stanley Durrleman. 2019. "Prediction of Future Cognitive Scores and Dementia Onset in Mild Cognitive Impairment Patients." In **Annual Meeting of the Organization for Human Brain Mapping – OHBM 2019**. <https://hal.inria.fr/hal-02098427>
2. **Manon Ansart**, Igor Koval, Anne Bertrand, Didier Dormont, and Stanley Durrleman. 2018. "Design of a Decision Support System for Predicting the Progression of Alzheimer's Disease." **Alzheimer's & Dementia** 14 (7) : P433. <https://doi.org/10.1016/j.jalz.2018.06.371>.
3. **Manon Ansart**, Stéphane Epelbaum, Olivier Colliot, Didier Dormont, Bruno Dubois, Harald Hampel, and Stanley Durrleman. 2017. "Selection of Amyloid Positive Pre-Symptomatic Subjects Using Automatic Analysis of Neuropsychological and MRI Data for Cost Effective Inclusion Procedures in Clinical Trials." In **The Journal of Prevention of Alzheimer's Disease**, 4 :P295. <https://hal.inria.fr/hal-02098434v2>.

TALKS AND POSTERS ---

1. Poster – Annual meeting of the Organization for Human Brain Mapping-OHBM, Rome, Italy, June 2019.
2. Poster – Alzheimer's Association International Conference, Chicago, USA, June 2018.
3. Oral presentation - Clinical Trials fro Alzheimer's Disease, Boston, USA, November 2017.

4. Oral presentation – Multimodal Learning for Clinical Decision Support workshop at the Medical Image Computing and Computer Assisted Intervention (MICCAI) conference. Quebec, Canada, September 2017.

SCIENTIFIC POPULARIZATION _____

1. Participation to the French scientific radio show "Le Club de la Tête au Carré" on France Inter, Paris France, 2019
2. Presentation in the Women in Machine Learning and Data Science (WiMLDS) meet-up, Paris France, 2019
3. Interview for a special edition of the French newspaper of record Liberation on artificial intelligence, Paris France, 2018
4. Fête de la science, Campus Jussieu, Paris, France, 2016
5. Salon Culture et Jeux Mathématiques, Paris France, 2016

Table des matières

Abstract	iii
Résumé	v
Scientific production	ix
Table des matières	xiii
Table des figures	xv
Liste des tableaux	xvii
List of Abbreviations	xix
Introduction	1
1 Predicting the Progression of Mild Cognitive Impairment Using Machine Learning : A Systematic and Quantitative Review	9
1.1 Introduction	10
1.2 Materials and Method	11
1.2.1 Selection process	11
1.2.2 Reading process	12
1.2.3 Quality check	12
1.2.4 Statistical analysis	14
1.3 Descriptive analysis	16
1.3.1 A recent trend	16
1.3.2 Features	16
1.3.3 Algorithm	17
1.3.4 Validation method	18
1.4 Performance analyses	18
1.4.1 Features	18
1.4.2 Cognition	19
1.4.3 Medical imaging and biomarkers	19
1.4.4 Combination of different imaging modalities	20
1.4.5 Longitudinal data	20

1.4.6	Algorithms	22
1.5	Design of the decision support system and methodological issues	22
1.5.1	Identified issues	22
1.5.1.1	Lack or misuse of test data	22
1.5.1.2	Performance as a function of data set size	22
1.5.1.3	Use of features of test subjects	24
1.5.1.4	Use of the diagnosis date	25
1.5.1.5	Choice of time-to-prediction	25
1.5.1.6	Problem formulation and data set choice	26
1.5.2	Proposed guidelines	27
1.6	Conclusion	27
2	Prediction of future cognitive scores and dementia onset in Mild Cognitive Impairment patients	29
2.1	Introduction	29
2.2	Materials and methods	31
2.2.1	Cross-sectional framework	31
2.2.1.1	Description	31
2.2.1.2	Inclusion of additional features	31
2.2.2	Longitudinal frameworks	32
2.2.2.1	Averaging approach	32
2.2.2.2	Temporal regression and stacking	33
2.2.2.3	Rate of change approach	33
2.2.3	Experimental setup	33
2.2.3.1	Data set	33
2.2.3.2	Validation procedure	34
2.2.4	TADPOLE challenge	34
2.3	Results	34
2.3.1	Cross-sectional framework	34
2.3.1.1	Proposed approach	34
2.3.1.2	Additional features	36
2.3.1.3	Building regression groups	36
2.3.2	Longitudinal frameworks	36
2.3.2.1	Averaging approach	36
2.3.2.2	Stacking approach	37
2.3.2.3	Rate of change approach	37
2.3.3	Prediction at different temporal horizons	38
2.3.4	TADPOLE challenge	39
2.4	Discussion	40

2.4.1	Cross-sectional experiments	40
2.4.2	Longitudinal frameworks	41
2.4.3	Interpretability	41
2.4.4	TADPOLE challenge	41
2.5	Conclusion	42
3	Reduction of Recruitment Costs in Preclinical AD Trials : Validation of Automatic Pre-Screening Algorithm for Brain Amyloidosis	45
3.1	Abstract	46
3.2	Introduction	46
3.2.1	Background	46
3.2.2	Related works	48
3.2.3	Contributions	48
3.3	Materials and Methods	49
3.3.1	Cohorts	49
3.3.2	Input Features	50
3.3.3	Algorithms	51
3.3.4	Performance Measures	51
3.4	Results	53
3.4.1	Algorithm and feature choice	53
3.4.1.1	Algorithm choice	53
3.4.1.2	Feature selection for cognitive variables	54
3.4.1.3	Use of MRI	55
3.4.2	Use of longitudinal measurements	55
3.4.3	Proposed method performance	56
3.4.3.1	Cost reduction	56
3.4.3.2	Age difference between groups	57
3.4.3.3	Training on a cohort and testing on a different one	58
3.4.3.4	Representativity of the selected population	59
3.4.4	Building larger cohorts	60
3.4.4.1	Pooling data sets	60
3.4.4.2	Effect of sample size	61
3.5	Discussion	61
3.5.1	Results of the experiments	61
3.5.1.1	Algorithm and feature choice	61
3.5.1.2	Method performance	63
3.5.1.3	Data set size	63
3.5.2	Comparison with existing methods	64
3.5.2.1	Univariate approaches	64

3.5.2.2	Other multivariate approaches	64
3.6	Conclusion	65
4	Use of psychotropic drugs throughout the course of Alzheimer’s disease : a large-scale study of French medical records	67
4.1	Introduction	67
4.2	Materials and methods	68
4.2.1	Cohort description	68
4.2.1.1	Description	68
4.2.1.2	Group definition	68
4.2.1.3	Patient overview	69
4.2.2	Studied treatments	69
4.2.3	Descriptive and predictive analysis of treatment history	70
4.2.3.1	Statistical analysis	70
4.2.4	Predictive model	71
4.3	Results	72
4.4	Discussion	73
4.4.1	Risk factors	73
4.4.2	Prediction	79
4.4.3	Management practices	80
4.4.4	Strengths and weaknesses of the study	80
4.5	Conclusion	81
	Conclusion & Perspectives	83
A	Supplementary materials for the systematic and quantitative review	87
A.1	Query	87
A.2	Selection process diagram	87
A.3	Reported items	87
A.4	Journals and conference proceedings	90
A.5	Information table	90
B	Supplementary materials for amyloidosis prediction	93
B.1	Computing R and S from the PPV and NPR	93
B.2	Difference of age in the 3 cohorts	94
B.3	Algorithm pseudo-code	94
C	Supplementary materials for the study of treatment prescriptions	97
C.1	Statistical analysis	97
C.1.1	Model description	97
C.1.2	Coefficient interpretation	98

C.1.3	Intercept of the non-AD group	98
C.1.4	Slope of the non-AD group	98
C.1.5	Intercept change for the AD group	98
C.1.6	Slope change for the AD group	99
C.1.7	Impact of diagnosis on the intercept	99
C.1.8	Impact of diagnosis on the slope	99
C.2	Predictive model	99
C.2.1	Performance measures	99
C.2.2	Results optimized for screening	100

Table des figures

1.1	Recent trends. (a) Evolution of number of article per year (in red) and of the number of individuals per article with time (in blue). (b) Evolution of the area under the ROC (receiver operating characteristic) curve (AUC) with time. The AUC of each article is represented by a dot. The AUC of articles published the same year is represented as box-plots. The plain line corresponds to the regression of the AUC against time	15
1.2	Evolution of the use of various algorithms with time. OPLS : orthogonal partial least square; SVM : support vector machine	17
1.3	Relationship between the AUC (area under the ROC curve) and the number of individuals. The black dotted lines represent the upper and lower limits.	23
1.4	Evolution of the performance with respect to the time to prediction. Box plots represent the accuracy of the articles using ADNI. The straight line represents the accuracy of predicting that all individuals remain MCI, corresponding to the proportion of MCI individuals in ADNI staying MCI at the corresponding follow-up time. The shaded area corresponds to the 90% confidence interval of this percentage.	24
2.1	Proposed approaches : A. cross-sectional framework ; B. longitudinal framework with 3 propositions : (a) averaging, (b) stacking, (c) rate of change.	32
2.2	Mean absolute error (MAE) of the prediction of the four cognitive scores at one year. The dots correspond to the mean performance and the lines to the 95% confidence interval	35
2.3	Comparison of our performance with the AUC of methods identified in the literature review for different times to prediction. Box plots correspond the performance of other methods, the whiskers going from the minimum to the maximum value. Triangles represent the mean AUC of our approach. AUC : Area Under the ROC Curve.	39

3.1 Selecting beta-amyloid ($A\beta$)⁺ individuals : current (left) and proposed (right) process 49

3.2 Example of ROC curve (left), S vs R curve (middle) and corresponding cost curve (right). The solid curve represents the mean performance, the dotted ones represent the standard deviation and the 2 black dots the points of minimal cost 52

3.3 Performance variations depending on the number of kept LASSO variables : mean Area Under the ROC curve (plain line) and standard deviation (dashed lines) 54

3.4 Histogram of the different features for the selected group (orange) and for the whole $A\beta$ ⁺ group (blue), for the ADNI-CN cohort 59

3.5 Proportion of runs with a significant difference between the groups for each feature, in each of the 3 cohorts 60

4.1 Positive and negative cases for the different scenarios of the predictive model 72

4.2 Proportion of treated subjects such as observed in the cohort and such as fitted by the model, for the AD group (in red) and the CN group (in green). The bar plots show to the number of subjects observed at the corresponding time points and how many of them where treated (solid colors) or not (transparent colors) (y-axis on the left). The dashed lines correspond to the probability of being treated at each time point for the different groups, such as fitted by the model (y-axis on the right). 74

4.3 Proportion of treated subjects such as observed in the cohort and such as fitted by the model, for the AD group (in red) and the MCI group (in orange). The bar plots show to the number of subjects observed at the corresponding time points and how many of them where treated (solid colors) or not (transparent colors) (y-axis on the left). The dashed lines correspond to the probability of being treated at each time point for the different groups, such as fitted by the model (y-axis on the right). 75

A.1 Diagram representing who the articles were selected 88

Liste des tableaux

1.1	Impact of features and algorithm. Benjamini-Hochberg procedure was applied to get corrected p-values. coeff. :coefficient, such as defined in Equations 1.1 and 1.2; MRI : magnetic resonance imaging; APOE : Apolipoprotein E; FDG : fluorodeoxyglucose; PET : positron emission tomography; CSF : cerebrospinal fluid; EEG : electroencephalography; MEG : magnetoencephalography; LDA : linear discriminant analysis; MKL : multiple kernel learning; OPLS : orthogonal partial least square; SVM : support vector machine . . .	21
2.1	Area Under the ROC curve (AUC) and balanced accuracy (bacc) obtained on the tested approaches, in the form mean \pm standard deviation. Significant differences between the proposed approach and the other method is shown as : *= significant at the 0.05 level; **= significant at the 0.01 level; ***= significant at the 0.001 level . .	35
2.2	Coefficients of the regression combining the cross-sectional prediction and the time linear regression in the stacking approach. Data are mean (standard deviation).	37
2.3	Performance of the cross-sectional approach for prediction at 1 year, 2 years, 3 years, in terms of Mean Absolute Error (MAE), Area Under the ROC curve (AUC) and balanced accuracy (bacc).	38
2.4	Results obtained on the TADPOLE challenge. mAUC : multiclass area under the receiver operating curve; bacc : balanced accuracy; MAE : mean absolute error.	40
3.1	Benchmark of algorithms. * = statistically significantly different from the Random Forest at the 0.05 level after Bonferroni correction for multiple comparisons. Data are : average Area Under the ROC curve (standard deviation). SVM = support vector machine; AdaLogReg = adaptive logistic regression; AdaBoost = adaptive boosting; ROC = receiver operating characteristic.	53
3.2	Results in different experimental conditions. Data are : average percentage of Area Under the ROC Curve (standard deviation). NA = Not Applicable	56

3.3	Comparison of the proposed method results with the estimated initial costs for recruiting K=100 amyloid positive individuals. AUC = Area Under the ROC curve; std = standard deviation.	57
3.4	Results using MRI variables, socio-demographic and genetic information on different data sets. Data are : average Area Under the ROC curve in % (std)	58
4.1	Cohort description. Data are mean (standard deviation). * = significant at the 0.05 level; ** = significant at the 0.01 level; *** = significant at the 0.001 level(two-sided t-test with Bonferroni correction for multiple comparison).	69
4.2	Odd ratios of prescription practices before and after diagnosis of AD as compared to MCI control group. Data are odd ratios (Standard deviations). * = significant at the 0.05 level; ** = significant at the 0.01 level; *** = significant at the 0.001 level (Bonferroni correction for multiple comparison was applied). AD : Alzheimer's disease, MCI : mild cognitive impairment.	76
4.3	Odd ratios of prescription practices before and after diagnosis of AD as compared to CN control group. Data are odd ratios (Standard deviations). * = significant at the 0.05 level; ** = significant at the 0.01 level; *** = significant at the 0.001 level (Bonferroni correction for multiple comparison was applied). AD : Alzheimer's disease, CN : cognitively normal.	77
4.4	Performance of the prediction of the presence of an AD diagnosis 5 or 10 years after a random visit for different groups of subjects. AUC = area under the receiver operating characteristic curve; J = Youden's J statistic; acc = accuracy; bacc = balanced accuracy; sen = sensitivity; spe = specificity.	77
A.1	Number of included articles published in each journal or conference proceedings. Only the journals with more than one included article are shown here. The articles taken into account are the one considered for analysis, and that use a large enough data set.	91
B.1	Age comparison between A β - and A β + individuals for the different cohorts. std = standard deviation.	94
C.1	CPerformance of the prediction of the presence of an AD diagnosis 5 or 10 years after a random visit for different groups of subjects. AUC = area under the receiver operating characteristic curve; acc = accuracy; bacc = balanced accuracy; sen = sensitivity; spe = specificity.100	

List of Abbreviations

Aβ	Amyloid beta protein
acc	Accuracy
AD	Alzheimer's disease
ADASCog	Alzheimer's disease assessment scale cognitive sub-scale
ADNI	Alzheimer's Disease Neuroimaging Initiative
ApoE	Apolipoprotein E
ApoE4	Allele 4 of the apolipoprotein E
AUC	Area under the receiver operating characteristic curve
bacc	Balanced accuracy
CDR	Clinical dementia rating
CN	Cognitively normal
CSF	Cerebrospinal fluid
FDG	¹⁸ F 2-fluoro-2-deoxy-D-glucose
J	Youden's J statistic
MAE	Mean Absolute Error
MCI	Mild cognitive impairment
MCIp	Mild cognitive impairment that progressed to Alzheimer's Disease
MCI_s	Stable mild cognitive impairment
MMSE	Mini-mental state examination
MRI	Magnetic resonance imaging
PET	Positron emission tomography
RAVLT	Rey auditory verbal learning test
sen	sensitivity
spe	specificity
SUV_r	Standardized uptake value ratio
SVM	Support vector machine
T1	T1-weighted magnetic resonance imaging

Introduction

Context

With an estimated 46.8 million people living with dementia in 2015 according to PRINCE, WIMO et al. (2015), this condition is becoming a global health issue. As the global population is aging, this number is expected to increase, with 9.9 million new cases each year. These rising numbers result in an important economic burden on our health care systems, representing up to 1.09% of global GDP. Alzheimer's Disease (AD), which is responsible for 60 to 80% of dementia cases (*What is Alzheimer's? 2019*), is the sixth leading cause of death in the United States according to the National Institute of Health.

Dementia is often diagnosed late in the disease process, and a large number of cases remain undiagnosed : according to the World Alzheimer Report of 2011, only 20 to 50% of dementia cases are diagnosed in high-income countries, and even fewer in low-income countries (PRINCE, BRYCE et FERRI, 2011). The identification of individuals who are the most at risk to develop AD is essential for the implementation of early therapeutic interventions and prevention measures. It allows patients with dementia to plan ahead their future care when they still can, and to get early treatment to stabilize their cognition and delay the onset of the symptoms. Identifying patients at the beginning of the disease course can also help research on dementia, by allowing the study of the disease mechanisms over larger and earlier time periods than today.

Several therapeutic hypotheses are currently tested regarding AD, however many of the recent clinical trials did not lead to satisfactory results. An hypothesis regarding these failures is that treatments are tested too late in the disease process, when cognitive damage has already occurred and cannot be reversed. The focus is therefore now shifting to earlier stages of the disease. Early identification of at risk subjects, as well as the identification of different groups of subject having a similar profile, can help select patients for these clinical trials. If treatments targeting pre-symptomatic or early symptomatic stages of AD proved to be effective, such tools could be used to identify the patients that could benefit from these treatments.

Machine learning is a branch of statistics that relies on the use of algorithms to identify patterns in a data set, and exploit these patterns to make prediction

regarding new data points. Machine learning methods can therefore be used to either model the typical evolution of the disease, or to make predictions regarding specific individuals, in order to select at-risk patients and to predict their evolution. It represents a great opportunity to make predictions that are tailored to each patient, thus paving the way for precision medicine.

In this thesis, we propose to apply machine learning techniques to build decision support systems for selecting individuals who are at risk of developing Alzheimer's disease, in both research and clinical settings.

Alzheimer's Disease

Alzheimer's disease is a progressive disease, with early signs that can be observed decades before diagnosis. The process leading to cognitive impairment and the ordering of biomarker changes has been notably hypothesized by JACK, KNOPMAN et al. (2010).

AD is defined by the presence of abnormal protein deposits of two types. First, the abnormal processing of the amyloid precursor protein leads to an aggregation of amyloid proteins, forming amyloid plaques in the brain. Second, hyperphosphorylated tau proteins form neurofibrillary tangles inside neurons. The presence of these two lesions is what defines AD neuropathologically and distinguishes it from other neurodegenerative diseases (JACK, BENNETT et al., 2018). The concentration of these two proteins can be measured in the cerebrospinal fluid (CSF). Amyloid concentration in the CSF decreases and tau/phosphorylated tau concentrations increase as their brain deposit increases. Amyloid plaques can also be visualized using positron emission tomography (PET) imaging, with Pittsburgh Compound B (PiB) or fluorine-18 (F-18) tracers (e.g. florbetapir and florbetaben). Research regarding the development of such tracers for tau imaging is currently ongoing. Tracers for tau imaging are now coming to the market.

The second biomarker change that can be expected in AD is a change in brain metabolism, measured using ^{18}F 2-fluoro-2-deoxy-D-glucose (FDG) PET imaging. These changes are followed by changes in brain structure, including brain atrophy, which are measured using structural MRI. Cognitive impairment is supposed to appear last, and is measured using cognitive and functional assessments, such as the Alzheimer's disease assessment scale cognitive sub-scale (ADASCog), the mini-mental state examination (MMSE), or the clinical dementia rating scale (CDR).

Studying these various biomarkers can help understand the disease and diagnose it. Machine learning techniques allow the identification of very early and

subtle changes, and can be used to detect small changes in brain images or cognitive assessments that are associated with the future onset of the disease. They can be used to estimate the current clinical status of an individual, and to predict how it is likely to evolve : individuals suffering from a Mild Cognitive Impairment (MCI) can evolve to be diagnosed with AD (MCI individuals progressing to AD, or MCIP) or not (MCI stable individuals, or MCIs), and identifying those who will can be especially useful for early diagnosis purposes.

Data sets

In order to facilitate the study of AD, the Alzheimer's Disease Neuroimaging Initiative (ADNI) study was created in 2004. This longitudinal, multicenter American study is still ongoing. Individuals are followed every 6 months to every year, and for each visit, a diagnosis (CN, MCI or AD) is given based on memory complaint and cognitive impairment. The study provides cognitive assessments, structural and functional imaging, amyloid and FDG PET scans and CSF measurements as well as genetic and socio-demographic information, for 1900 individuals to date. This data set is widely used in AD research, and has played an important role in the creation of a large number of automatic methods.

Data bases created for research purposes provide intensive tests, performed on on a large number of visits in a short period of time. They are ideal for training models, however they do not reflect the daily clinical practice. On the other hand, databases of medical records better reflect the current clinical practice, and the associated challenges : patients undergo fewer tests, they are not necessarily observed as frequently, and medical imaging is performed only on targeted patients with cognitive impairments. Cegedim, for example, gathers data from patients followed by general practitioners and specialists all over France. Such electronic health record data bases represent a good opportunity for the development of automatic methods that aim at being integrated in the clinical workflow.

Related work

Automatic diagnosis

A large number of existing articles using machine learning for AD pertains to automatic diagnosis : predicting the current clinical status of an individual, thus distinguishing cognitively normal (CN) subjects, MCI subjects and AD subjects, mostly based on imaging features. As cognitive measurements are often used to establish the diagnosis, they are not included as inputs in these methods.

A large number of methods work on features extracted from MRI. We observe three main types of feature extraction for studying structural changes : creation of density maps of white matter (WM), grey matter (GM), and cerebrospinal fluid (CSF); study of the cortical surface and its thickness; usage of pre-defined brain regions (RATHORE et al., 2017). CUINGNET et al. (2011) compared the results obtained using density maps, cortical thickness and hippocampus volume or shape, and found that for the classification of AD versus CN, the two whole brain methods perform well, whereas using the hippocampus yields a lower performance. Other methods focus on the use of FDG-PET, such as K. R. GRAY et al. (2012), which reaches an accuracy of 88 % for distinguishing between AD and CN subjects. Lastly, the combination of different modalities which can complement each other have been the focus of different methods, such as KIM et LEE (2018), which combines MRI, FDG-PET and CSF measurements.

Several reviews of the automatic diagnosis of AD have been proposed. HALLER, LOVBLAD et GIANNAKOPOULOS (2011) focuses on methods performing automatic diagnosis using MRI features, and discuss the difference between classification methods, which perform a prediction for each individual, and the study of group differences. FALAHATI, WESTMAN et SIMMONS (2014) studies methods performing automatic diagnosis based on MRI features, but also PET imaging and CSF measurements, and present the main types of imaging feature extraction used in these methods. RATHORE et al. (2017) propose a review of methods based on neuroimaging, as does SARICA, CERASA et QUATTRONE (2017), with a focus on random forest classifiers. Lastly, ARBABSHIRANI et al. (2017) reviews methods performing individual prediction based on neuroimaging for a range of brain disorders, and offers a broader view of the opportunities these methods offer, as well as the issues they raise.

The performance of such methods is very high, but one could question their use as a decision support system in clinical practice. They aim at reproducing a diagnosis made by a clinician, which can be easily obtained. An interesting use of these methods lies in their interpretation. They show that the input modalities contain information that allows to distinguish almost perfectly AD and CN individuals, showing that the metabolic or structural changes induces by AD are important. Method offering a visualization of the parts of the image which are involved in the prediction are useful to identify the brain regions are the most impacted. LEANDROU et al. (2018) gives a complete overview of such methods and their main findings.

Prediction of future diagnosis

If predicting the clinical status that would be given by a clinician at the current time point is not an interesting task for a clinical decision support system, predicting the future diagnosis that could be given to a patient can be useful. In particular, identify the MCI subjects would are likely to develop AD in the future is an interesting task, and a large number of automatic methods have been proposed.

Some of the proposed methods, especially early on, are extensions of methods that have been trained at distinguishing AD individuals from CN individuals, and that are then applied to MCI individuals. MCI patients that are labeled as CN by the classifiers are expected to stay MCI, whereas those who are labeled as AD are expected to be diagnosed with AD later on (NHO et al., 2010; CHINCARINI et al., 2011; DAVATZIKOS et al., 2011; CUI et al., 2011; WESTMAN, MUEHLBOECK et SIMMONS, 2012; DUKART, SAMBATARO et BERTOLINO, 2015; RETICO et al., 2015; JUNWEI DING et QIU HUANG, 2017; CHOI et JIN, 2018). As for automatic diagnosis, a large number of previous methods have focused on the use of neuroimaging, and on MRI in particular. Several methods focus solely on MRI features, that can be voxel based (BEHESHTI, DEMIREL et MATSUDA, 2017; TONG et al., 2017; SABUNCU, 2013; YE, POHL et DAVATZIKOS, 2011) or region based, on the whole brain (MINHAS et al., 2018; RETICO et al., 2015; RAAMANA et al., 2015; WESTMAN, AGUILAR et al., 2013) or specific regions (KAUPPI et al., 2018; ARDEKANI et al., 2017; TANPITUKPONGSE et al., 2017; HALL et al., 2015; ESKILDSEN et al., 2015; CHINCARINI et al., 2011). CUINGNET et al. (2011) compares the performance of these types of features to predict the evolution of MCI at 18 months, but shows that the performance is not better than chance. GÓMEZ-SANCHO, TOHKA et GÓMEZ-VERDEJO (2018) shows that for prediction at 3 years, regional features across the brain perform better than voxel features or using the hippocampus only, although the difference is not significant.

Other methods use neuroimaging more broadly, and integrate features of various modalities (VIVAR et al., 2018; SI, YAKUSHEV et J. LI, 2017; JUNWEI DING et QIU HUANG, 2017; OTA et al., 2015; DUKART, SAMBATARO et BERTOLINO, 2015; SAMPER-GONZALEZ et al., 2019). These approaches may include the current cognitive assessments to predict the future diagnosis. They result in general in an increased performance (MINHAS et al., 2018; KAUPPI et al., 2018; ARDEKANI et al., 2017; TONG et al., 2017; MUBEEN et al., 2017; KOROLEV et al., 2016; LEI et al., 2016; SHAFFER et al., 2013).

Another point that differentiates the methods proposed to automatically predict the evolution of MCI is the choice of temporal horizon. Several methods distinguish MCIs from MCIp individuals with no fixed time to prediction, so for each

individual the considered time interval can be different (VECCHIO et al., 2018; LEI et al., 2016; HALL et al., 2015; DOYLE et al., 2014). Other methods make predictions at a specific time interval, ranging from 6 months to several years.

Contributions

Several automatic methods have been proposed to predict the evolution of the clinical status of subjects with a mild cognitive impairment. These methods vary greatly in terms of feature types, algorithm and test data set. We identified several reviews regarding the use of machine learning in Alzheimer's disease, but none of them focused specifically on the progression of MCI subjects to AD, and they do not provide a quantitative comparison of these articles. We propose a systematic and quantitative review of these methods : we study 172 articles, and for each one we take note of 36 key elements regarding the method, the input features and the test framework in order to compare them. We thus identify current trends in the domain, and study the impact of various methodological elements on the performance of the methods. We also study the usability of such methods as decision-support systems in clinical practice, and recommend several key-points that should be taken into accounts when building such systems.

Secondly, we propose our own method to predict the evolution on mild cognitive impairment and test it on the ADNI cohort. We propose a method composed of two parts : in a first part, we predict the evolution of cognitive scores, using previous measures of cognition, sub-cortical brain volumes and socio-demographic information. In a second part, we use this estimation of future patient state to predict the corresponding clinical status. We believe that this two-part prediction is easier to interpret for clinicians, is thus more likely to be adopted in clinical practice. In order to ensure our prediction is as accurate as possible, we study the impact of including additional features and longitudinal information.

Thirdly, we propose a tool to help select subjects for clinical trials at a lower cost. As amyloid deposit is one of the first signs of AD, a dominant hypothesis is that the formation of amyloid plaques triggers the cascade of events leading to AD. Several potential AD treatments thus target this protein, with the hope that clearing it or stopping its formation would stop this cascade. Clinical trials testing such treatments require to form a cohort of individuals for whom amyloid deposit has already started but who don't have any cognitive impairment yet, in order to target the earliest stages of the disease. We thus propose an automatic method to identify a group of individuals at risk of having these plaques based on specific signatures in cognitive and/or imaging data, thus leaving confirmatory PET scans or lumbar puncture for a smaller set of individuals. We believe that this tool, by

lowering the cost of recruitment of amyloid targeting clinical trials, could make the creation of such trials easier and hence facilitate therapeutic research.

Most of the articles using machine learning for identification of patients at risk of developing AD do so using research cohorts, which do not reflect the complexity of clinical practice. In a last study, we propose to use clinical data in order to build a decision support system which could be used in clinical routine. Using Cegedim, a database of medical records from general practitioners in the French health care system, we study the longitudinal evolution of treatment prescription of AD patients, and compare it to other cohorts of MCI or control subjects. We then build a decision-support system to identify patient who will be diagnosed with AD in the next 5 or 10 years based on their treatment history.

Chapitre 1

Predicting the Progression of Mild Cognitive Impairment Using Machine Learning : A Systematic and Quantitative Review

This chapter has been submitted to the Medical Image Analysis journal, as :

— **Manon Ansart**, Stéphane Epelbaum, Giulia Bassignana, Alexandre Bône, Simona Botani, Tiziana Cattai, Raphaël Couronné, Johann Faouzi, Igor Koval, Maxime Louis, Elina Thibeau-Sutre, Junhao Wen, Adam Wild, Ninon Burgos, Didier Dormont, Olivier Colliot, Stanley Durrleman. “Predicting the Progression of Mild Cognitive Impairment Using Machine Learning : A Systematic and Quantitative Review.”

Abstract

Context Automatically predicting if a subject with Mild Cognitive Impairment (MCI) is going to progress to Alzheimer’s disease (AD) dementia in the coming years is a relevant question regarding clinical practice and trial inclusion alike. A large number of articles have been published, with a wide range of algorithms, input variables, data sets and experimental designs. It is unclear which of these factors are determinant for the prediction, and affect the predictive performance that can be expected in clinical practice. We performed a systematic review of studies focusing on the automatic prediction of the progression of MCI to AD dementia. We systematically and statistically studied the influence of different factors on predictive performance.

Method The review included 172 articles, 93 of which were published after 2014. 234 experiments were extracted from these articles. For each of them, we reported the used data set, the feature types (defining 10 categories), the algorithm type (defining 12 categories), performance and potential methodological issues. The impact of the features and algorithm on the performance was evaluated using t-tests on the coefficients of mixed effect linear regressions.

Results We found that using cognitive, fluorodeoxyglucose-positron emission tomography or potentially electroencephalography and magnetoencephalography variables significantly improves predictive performance compared to not including them ($p=0.046$, 0.009 and 0.003 respectively), whereas including T1 magnetic resonance imaging, amyloid positron emission tomography or cerebrospinal fluid AD biomarkers does not show a significant effect. On the other hand, the algorithm used in the method does not have a significant impact on performance. We identified several methodological issues. Major issues, found in 23.5% of studies, include the absence of a test set, or its use for feature selection or parameter tuning. Other issues, found in 15.0% of studies, pertain to the usability of the method in clinical practice. We also highlight that short-term predictions are likely not to be better than predicting that subjects stay stable over time. Finally, we highlight possible biases in publications that tend not to publish methods with poor performance on large data sets, which may be censored as negative results.

Conclusion Using machine learning to predict MCI to AD dementia progression is a promising and dynamic field. Among the most predictive modalities, cognitive scores are the cheapest and less invasive, as compared to imaging. The good performance they offer question the wide use of imaging for predicting diagnosis evolution, and call for further exploring fine cognitive assessments. Issues identified in the studies highlight the importance of establishing good practices and guidelines for the use of machine learning as a decision support system in clinical practice.

1.1 Introduction

The early diagnosis of Alzheimer's disease (AD) is crucial for patient care and treatment. Machine learning algorithms have been used to perform automatic diagnosis and predict the current clinical status at an individual level, mainly in research cohorts. Individuals suffering from mild cognitive impairment (MCI) are however likely to have a change of clinical status in the coming years, and to be diagnosed with AD or another form of dementia. Distinguishing between the MCI

individuals that will remain MCI (MCI stable, or sMCI) from those who will progress to AD (pMCI) is an important task, that can allow for the early care and treatment of pMCI patients. In this article, we will review methods that have been proposed to automatically predict if an MCI patient will develop AD dementia in the future by performing a careful reading of published articles, and compare them through a quantitative analysis.

The application of machine learning to precision medicine is an emerging field, at the cross roads of different disciplines, such as computer science, radiology or neurology. Researchers working on the topic usually come from one field or the other, and therefore do not have all the skills that are necessary to design methods that would be efficient and following machine learning best practices, while being understandable and useful to clinicians.

Reviews of the automatic prediction of the current clinical diagnosis in the context of AD have already been published, but none specifically target the prediction of progression from MCI to AD dementia. They focus on the use of magnetic resonance imaging (MRI) (FALAHATI, WESTMAN et SIMMONS, 2014; LEANDROU et al., 2018), or of neuroimaging data more broadly (RATHORE et al., 2017; ARBABSHIRANI et al., 2017; HALLER, LOVBLAD et GIANNAKOPOULOS, 2011; SARICA, CERASA et QUATTRONE, 2017). Several of them are systematic reviews such as ARBABSHIRANI et al. (2017) with 112 studies on AD, RATHORE et al. (2017) with 81 studies, FALAHATI, WESTMAN et SIMMONS (2014) with 50 studies and SARICA, CERASA et QUATTRONE (2017) with 12 studies. They often gather the findings of each individual article and compare them, but no quantitative analysis of performance is proposed.

We propose here to perform a systematic and quantitative review of studies predicting the evolution of clinical diagnosis in individuals with MCI. We will report different quantitative and qualitative characteristics of the proposed method such as the sample size, type of algorithm, reported accuracy, identification of possible issues. We will then analyze this data to identify the characteristics which impact performance the most, and propose a list of recommendations to ensure that the performance is well estimated, and that the algorithm would have the best chance to be useful in clinical practice.

1.2 Materials and Method

1.2.1 Selection process

The query used to find the relevant articles was composed of 4 parts :

1. As we study the progression from MCI to AD, the words MCI and AD should be present in the abstract;
2. We removed the articles predicting only the current diagnosis by ensuring the words “prediction” and “progression” or associated terms are present in the abstract;
3. A performance measure should be mentioned;
4. A machine learning algorithm or classification related key-word should be in the abstract. This fourth part ensures the selected articles make individual predictions and reduces the presence of group analyzes.

The full query can be found in [A.1](#). Running it on Scopus on the 13th of December 2018 resulted in 330 articles. The abstracts were read to remove irrelevant articles, including studies of the progression of cognitively normal individuals to MCI, automatic diagnosis methods, review articles and group analyses. After this selection 206 articles were identified. As this first selection was quite conservative, 34 additional articles were removed from the selection for similar reasons during the reading process, leaving 172 studied articles. The selection process is described in [Figure A.1](#) in [A.2](#).

1.2.2 Reading process

For each study, the number of individuals was first assessed and noted. Only studies including more than 30 sMCI and 30 pMCI (111 articles) were then fully read, as we consider that experience using less than 30 individuals cannot provide robust estimates of performance. Articles with less than 30 individuals in each category were still considered when studying the evolution of the number of articles with time, and of the number of individuals per article with time. The studies including enough individuals were then analyzed by one of the 19 readers participating in this review, and a global check was performed by one author (MA) to ensure homogeneity. 36 items, of which a list is available in [A.3](#), were reported for each study, including the used features, the cohort, the method (time to prediction, algorithm, feature selection, feature processing), the evaluation framework and the performance measures, as well as identified biases in the method. When several experiments were available in an article, they were all reported in the table. A total of 234 experiments was thus studied.

1.2.3 Quality check

Several methodological issues were identified during the reading process. This list of issues was not previously defined, it has been established as issues were

encountered in the various studies. We identified the following list of issues :

- Lack of a test data set : use of the same data set for training and testing the algorithm, without splitting the data set or using any kind of cross-validation method. The performance computed this way is the training performance, whereas a test performance, computed on a different set of individuals, is necessary to measure the performance that could be obtained on any other data set (i.e. generalizability of the method) .
- Automatic feature selection performed on the whole data set. When a large number of features is available, automatic feature selection can be performed in order to identify the most relevant features and use them as input. A variety of automatic algorithms exist to do this. Some studies performed this automatic feature selection on the whole data set, before splitting it into a training and a test set or performing cross-validation. An example of this issue is, first, using t-tests to identify features that best separate pMCI from sMCI, using the whole data set, then splitting the data set into a training and a test set, to respectively train the classification algorithm and evaluate its performance. In this example, the individuals from the test set have been used to perform the automatic feature selection and choose the most relevant features. This is an issue, as individuals in the test set should be used for performance evaluation only.
- Other data-leakage. More broadly, data leakage is the use of data from the test set outside of performance evaluation. Using the test data set for parameter tuning, or for choosing the best data set out of a large number of experiments, are two common examples of data leakage.
- Feature embedding performed on the whole data set. Feature embedding (for example principal components analysis) transforms the input features into a lower-dimension feature space. It is often used to reduce the input dimension when many features are available, but it does not use the individual labels (sMCI/pMCI) to do so, as feature selection often does. This issue is therefore similar to performing feature selection on the whole data set, except that only the features of the test individuals are used, and not their labels.
- Use of the date of AD diagnosis to select the input visit of pMCI individuals. An example of this issue is using the visit 3 years before progression to AD for pMCI subjects, and the first available visit for sMCI subjects, to predict the progression to AD at 3 years, even for testing the method. In this case, the date of progression to AD of the individuals of the test set was used to select the input visit, which is not possible in clinical practice, as the date of progression is not known.

Other methodological issues, not belonging to these categories, were also reported, such as incompatibility between different reported measures. The articles in which at least one of these issues was identified were not used when analyzing the performance of the methods and the method characteristics impacting them.

1.2.4 Statistical analysis

We studied the impact of various method characteristics (input features, algorithm...) on the performance of the classification task, separating sMCI from pMCI individuals. Several experiments were reported for each article, so we had to account for the dependency between experiments coming from the same article. In order to do so, we used linear mixed-effects models with a random effect on the article, and tested if the considered characteristics had a significant impact by performing a two-sided t-test on the corresponding regression coefficient. Only the characteristics found in more than one article with an associated performance measure were taken into account. Unless stated otherwise, the performance measure used for testing is the area under the receiver operating characteristic (ROC) curve (AUC), experiments with no reported AUC were therefore not taken into account in these tests. When testing the impact of various characteristics at the same time, conditionally to each other (e.g. among all input features, which ones have an impact on the performance when taking the other features into account), we performed a linear mixed effect regression with all these characteristics as input. Concerning the input features, d being the number of features :

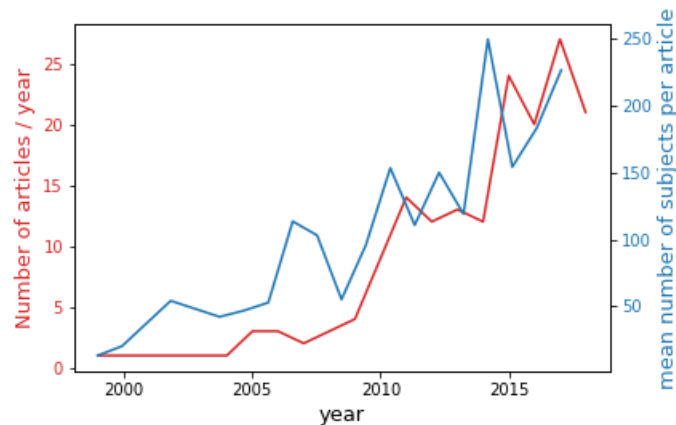
$$AUC = \alpha_1 * feature_1 + \dots + \alpha_d * feature_d + \beta + \beta_{article} \quad (1.1)$$

When testing the impact of different characteristics independently (e.g. for each algorithm, the effect of using this specific algorithm or any other), an individual linear mixed effect regression was performed for each one separately :

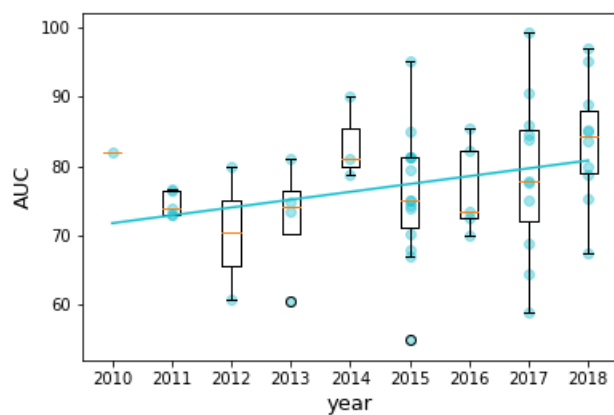
$$AUC = \alpha_i * algo_i + \beta + \beta_{article} \quad (1.2)$$

for all i , i being the algorithm number.

In both cases, a two-sided t-test was performed on α to test the significance of each coefficient. The p-values corrected for multiple comparisons were obtained by using the Benjamini-Hochberg procedure.



(A) Evolution of number of article per year and of the number of individuals per article



(B) Evolution of the AUC with time

FIGURE 1.1 – Recent trends. (a) Evolution of number of article per year (in red) and of the number of individuals per article with time (in blue). (b) Evolution of the area under the ROC (receiver operating characteristic) curve (AUC) with time. The AUC of each article is represented by a dot. The AUC of articles published the same year is represented as box-plots. The plain line corresponds to the regression of the AUC against time

1.3 Descriptive analysis

1.3.1 A recent trend

We observe from Figure 1.1a that the number of articles published each year on the prediction of the progression of MCI to AD dementia has been steadily increasing since 2010.

Figure 1.1a also shows that the number of individuals used for the experiments is increasing over time ($p=10^{-5}$). 84.6% of articles used data of the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. Starting in 2004, this multicenter longitudinal study provides multiple modalities for the early detection of AD. As the recruitment of this largely used cohort is still ongoing, it is not surprising to see the number of included individuals increasing over the years. Studies often select individuals with a minimal follow-up time, of 3 years for example, and over the years more and more MCI individuals from the ADNI cohort fulfill these criteria, so more individuals can be included.

As shown in Figure 1.1b, the reported AUC are also increasing over time ($p=0.045$), which can have multiple explanations. First, as new studies often compare their performance with those of previous methods, they tend to be published only when the obtained results seem competitive compared to previous ones. A more optimistic interpretation would be that algorithms tend to improve, and that newly available features might have a better predictive power. It has also been shown (ANSART, EPELBAUM, GAGLIARDI, COLLIOT, DORMONT, DUBOIS et al., 2019; DOMINGOS, 2012) that having a larger data set leads to a higher performance, so there may be a link between the increase in data set size and the increase in performance.

1.3.2 Features

T1 MRI, cognition and socio-demographic features are used in respectively 69.2%, 43.2% and 33.8% of experiments. On the other hand, fluorodeoxyglucose (FDG) positron emission tomography (PET), APOE and cerebrospinal fluid (CSF) AD biomarkers are used in 15 to 20% of experiments, and the other studied features (white matter hyper-intensities, electroencephalography (EEG), magnetoencephalography (MEG), PET amyloid, amyloid binary status without considering the PET or CSF value, diffusion tensor imaging (DTI) and PET Tau) are used in less than 10% of experiments. No study using functional MRI has been identified.

Studies using T1 MRI mainly use selected regions of interest (46.8%), whereas 34.7% use the whole brain, separated into regions of interest, and 18.5% use voxel

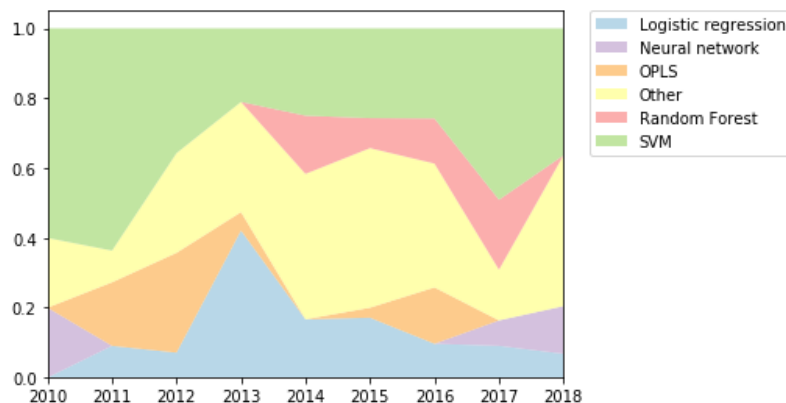


FIGURE 1.2 – Evolution of the use of various algorithms with time. OPLS : orthogonal partial least square ; SVM : support vector machine

features. Studies using neuro-psychological tests mainly use aggregated tests evaluating multiple domains of cognition (51.2% of them), and 37.4% of them combine aggregated tests with domain-specific ones. Seven experiments use new or home-made cognitive tests. 35.7% of experiments use only T1 MRI (apart from socio-demographic features), and 15% use cognition only.

The prevalence of T1 MRI does not seem surprising, as researchers working on automatic diagnosis often come from the medical imaging community, and T1 MRI is the most widely available modality. The prevalence of the imaging community can also explain the choice of cognitive features, and why more detailed and targeted cognitive tests are not used as much as more general and more well-known ones.

1.3.3 Algorithm

Support vector machines (SVM) and logistic regressions are the most used algorithms, being used in respectively 34.5% and 15.0% of experiments. Other algorithms are used in less than 10% of cases. Figure 1.2 shows the evolution of the algorithm use over time.

The high proportion of methods using an SVM has already been shown for the prediction of the current diagnosis in FALAHATI, WESTMAN et SIMMONS (2014) and RATHORE et al. (2017), it is therefore not surprising that this algorithm is also commonly used for the prediction of future diagnosis. The predominance of SVM and logistic regression still seems surprising, as more recent algorithms are more popular nowadays. We see for example that random forests started being used around 2014, but the proportion of methods using this algorithm, even recently, stays low compared to the proportion of methods using an SVM. Neural networks started being used during the last two years, as it can be seen in Figure 1.2, and we

can assume the phenomenon has been too recent to be visible just yet in the field. Overall, even if the proportion of SVM has been decreasing until 2013, the field has not been so prompt to use new algorithms as one could have expected.

1.3.4 Validation method

For evaluating their performance, 29.1 % of experiments use a 10-fold, and 12.8% use a k-fold with k different from 10. Leave-one individual out is also quite popular, being used in 17.5% of cases. We noted that 7.3% of experiments were trained and tested on the same individuals, and 7.3% train the method on a first cohort and test it on a different one.

It should be kept in mind when comparing the performance of different studies that the cross-validation methods can impact the performance. Using a larger training set and smaller test set is more favorable, hence the same method might result in a better performance when evaluated using a leave-one out validation than using a 10-fold validation, as shown in W. LIN et al. (2018). Bias and variance also vary across validation methods (EFRON, 1983).

1.4 Performance analyses

1.4.1 Features

We measured the impact on the AUC of each feature compared to the others by using a linear mixed-effect model including all features used in more than one article. The results are presented in the first part of Table 1.1, and show that the performance is significantly better when using cognition ($p = 0.046$), FDG PET ($p=0.009$) or EEG and MEG ($p=0.003$).

We also considered the impact of using each feature alone compared to a combination of them, by testing each feature independently using a linear mixed effect regression. We only tested the impact of the features that were used alone (or in combination with socio-demographic features) more than once with an associated AUC, and that had been combined with other features more than once, that is T1 MRI, cognition, and FDG PET. It is significantly better to combine T1 MRI with other features than to use it solely ($p = 0.009$, coefficient = 5.5). The effect is not significant for cognition ($p=0.19$, coefficient=3.0) and FDG PET ($p=0.38$, coefficient = -6.1).

We distinguished between global neuro-psychological tests, domain-targeted tests and newly proposed tests. We measured the impact of the type of test on the AUC by performing independent regressions for each category. Experiences

using a domain-specific test had a significantly greater AUC than those that did not ($p=0.023$, coefficient = 5.0), whereas the effect was not significant for the other two categories ($p > 0.1$). We tested the impact on the AUC of using longitudinal data (repeated visits as input), and of combining images of different modalities, and both were not significant ($p > 0.2$)

1.4.2 Cognition

Cognitive variables can be easily collected in clinical routine, at a low cost, and they are proven to increase the performance of the methods, so their use should be encouraged. This finding is consistent with comparisons performed in several studies. MINHAS et al. (2018), KAUPPI et al. (2018), ARDEKANI et al. (2017), TONG et al. (2017), GAVIDIA-BOVADILLA et al. (2017), MORADI et al. (2015), HALL et al. (2015) et FLEISHER et al. (2008) showed that using cognition and T1 MRI performed better than using T1 MRI only. DUKART, SAMBATARO et BERTOLINO (2015), CUI et al. (2011), THUNG et al. (2018) et Y. LI et al. (2018) showed that adding cognition to other modalities also improved the results.

More surprisingly, we showed that using other modalities does not significantly improve the results compared to using cognition only. Although FLEISHER et al. (2008) shows that using T1 MRI in addition to cognition does not improve the performance compared to using cognition only, several studies show the opposite on various modalities (SAMPER-GONZALEZ et al., 2019; MORADI et al., 2015; ARDEKANI et al., 2017; Y. LI et al., 2018; KAUPPI et al., 2018). However, when taking all studies into account, it appears that the improvement one gains by including other modalities along with cognitive variables is not significant. As the cost of collecting cognitive variables compared to performing an MRI or a FDG PET is quite low, the non-significant improvement in performance might not be worth the cost and logistics of collecting data from other modalities specifically to address this question. Methods focusing on cognition only, such as proposed by JOHNSON et al. (2014), should therefore be further explored. Such methods should include domain-specific cognitive scores, which have shown to increase the performance.

1.4.3 Medical imaging and biomarkers

Imaging modalities are not as widely available as cognitive feature, but they can represent a good opportunity to better understand the disease process by showing the changes that appear before the individuals progress to AD dementia. Among the used imaging modalities, we showed that using FDG PET leads to a

better performance. Similar observations have been made by SAMPER-GONZÁLEZ et al. (2018). PET images could therefore represent a better alternative for the imaging community than T1 MRI, which does not significantly improve the results and should not be used alone as it leads to lower results. Changes in FDG PET appear earlier in the AD process than changes in structural MRI (JACK, KNOPMAN et al., 2010), therefore these changes might already be visible in MCI individuals several years before their progression to AD, which can explain why FDG PET is more predictive of this progression.

No method using Tau PET has been identified in this review. This new modality should also be affected early in the disease process, and could therefore represent great hopes for the imaging community. However, surprisingly, Amyloid PET or CSF value, which is also one of the earliest markers, did not have a significant impact on the prediction performance.

The use of EEG or MEG had a significant impact on the performance. However, only six experiments use these features, it is therefore difficult to conclude if this effect is real, and if it is not due to methodological issues that have not been identified during the quality check.

1.4.4 Combination of different imaging modalities

Multimodality has been put forward in the reviews of AD classification (RATHORE et al., 2017; FALAHATI, WESTMAN et SIMMONS, 2014; ARBABSHIRANI et al., 2017). As different imaging modalities correspond to various stages of the AD process, combining them could give a more complete overview of each individual. However, we did not find the impact of the use of multimodality to be significant. This result is not surprising, as the most combined modalities are MRI and FDG PET (19 out of 35 experiments using multimodality), and we showed that including other features does not lead to a significant increase in performance compared to using FDG PET alone. In addition, the cost of collecting images of different modalities for each patient is not small, and should not be neglected when using such approaches.

1.4.5 Longitudinal data

In a similar manner, longitudinal data could give a better view of the evolution of the patient, and hence be more predictive of the progression to AD than cross-sectional data. Nonetheless, we did not find the use of longitudinal data to have a significant effect on the performance. Similar findings are reported in AKSMAN (2017) for the classification of AD and in SCHUSTER et al. (2015) for progressive diseases in general.

Feature or algorithm	coeff.	p-value	corrected p-value	number of exp.
T1 MRI	2.217	0.22	0.38	103
Neuro-psychological tests	3.934	0.046	0.11	64
socio-demographic	0.652	0.83	0.83	59
APOE	4.612	0.092	0.18	35
FDG PET	6.768	0.0092	0.037	29
CSF	2.232	0.38	0.41	26
Others	3.12	0.28	0.4	18
EEG/MEG	16.573	0.0025	0.015	6
PET Amyloid	7.743	0.3	0.4	6
White matter hyper-intensities	-5.18	0.36	0.41	5
SVM	-4.8	0.061	0.24	35
Logistic regression	0.8	0.812	0.93	15
Random Forest	4.1	0.166	0.5	13
MKL	-0.3	0.950	0.95	10
Other	0.8	0.851	0.93	7
Bayes	5.4	0.271	0.65	6
Linear regression	-5.2	0.434	0.74	6
Neural network	10.1	0.010	0.06	6
OPLS	-15.5	0.003	0.04	6
Survival analysis	2.0	0.810	0.93	6
Threshold	1.1	0.791	0.93	6
LDA	-6.3	0.325	0.65	5

TABLE 1.1 – Impact of features and algorithm. Benjamini-Hochberg procedure was applied to get corrected p-values. coeff. :coefficient, such as defined in Equations 1.1 and 1.2; MRI : magnetic resonance imaging; APOE : Apolipoprotein E; FDG : fluorodeoxyglucose; PET : positron emission tomography; CSF : cerebrospinal fluid; EEG : electroencephalography; MEG : magnetoencephalography; LDA : linear discriminant analysis; MKL : multiple kernel learning; OPLS : orthogonal partial least square; SVM : support vector machine

1.4.6 Algorithms

We studied the impact of the algorithms on the AUC, by using an independent linear mixed effect model on each algorithm. The results, presented in the second part of Table 1.1, show that the orthogonal partial least square (OPLS) algorithm performs significantly worse than others ($p=0.003$), whereas neural networks perform significantly better ($p=0.01$).

Only six experiments have been performed using each of these algorithms, so an unidentified methodological issue in one of them could greatly impact these results. As neural networks have a large number of parameters, which are often tuned manually using the test error, we found that experiments using this algorithm have high proportion of data leakage. This is consistent with the findings of WEN et al. (2019), a literature review conducted on the use of deep learning for AD classification. No conclusion regarding the impact of the classification algorithm can therefore be drawn from our results, which might be explained by the variety of algorithms, and hence the small sample size for each of them.

1.5 Design of the decision support system and methodological issues

1.5.1 Identified issues

1.5.1.1 Lack or misuse of test data

The lack of a test data set is observed in 7.3% of experiments. In 16% of articles using feature selection, it is performed on the whole data set, and 8% of articles do not describe this step well enough to draw conclusions. Other data leakage (use of the test set for decision making) is identified in 8% of experiments, and is unclear for 4%.

Overall, 26.5% of articles use the test set in the training process, to train the algorithm, choose the features or tune the parameters. This issue, and in particular performing feature selection on the whole data set, has also been pointed out by ARBABSHIRANI et al. (2017) in the context of brain disorder prediction.

1.5.1.2 Performance as a function of data set size

We plot the AUC against the number of individuals for each experiment in Figure 1.3, with the colored dots representing experiments with identified issues. The colored dots show that there is a higher prevalence of studies with identified issues among high-performance studies : a methodological issue has been

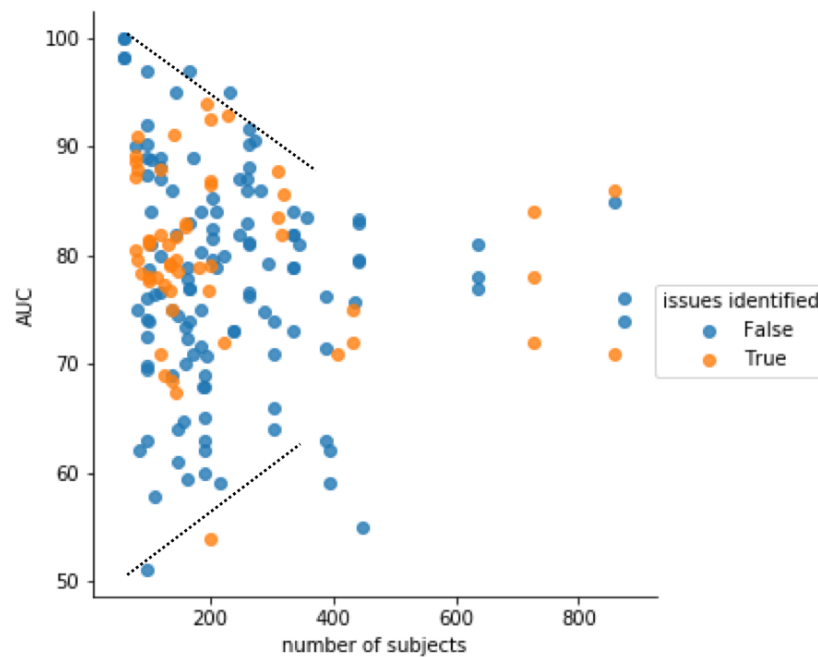


FIGURE 1.3 – Relationship between the AUC (area under the ROC curve) and the number of individuals. The black dotted lines represent the upper and lower limits.

identified in 18.5% of experiments with an AUC below 75%, whereas this proportion rises to 36.4% for experiments with an AUC of 75% or higher (significant difference, with $p = 0.006$). We can observe an upper-limit (shown in dashed line) decreasing when the number of individuals increases, suggesting that high-performance achieved with a small number of subjects might be due to overfitting. This phenomenon has already been identified by ARBABSHIRANI et al. (2017). A lower limit is also visible, with the AUC increasing with the number of individuals. This may reflect the fact that, on average, methods generalize better when correctly trained on larger data sets. But it might also suggest that it is harder to publish a method with a relatively low performance if it has been trained on a large number of subjects, such a paper being then considered as reporting a negative result. Within papers also, authors tend to focus on their best performing method, and rarely explain what they learned to achieve this. As the number of subjects increases, the two lines seem to converge to an AUC of about 75%, which might represent the true performance for current state-of-the-art methods.

Figure 1.3 seems to highlight possible unconscious biases in the publications of scientific results in this field. It might be considered more acceptable to publish high-performance methods with small sample size than a low-performance method with large sample size. First, we think that low-performance methods trained on large sample size should be published also, as it is important for the field to understand what works and also what does not. In particular, we think that

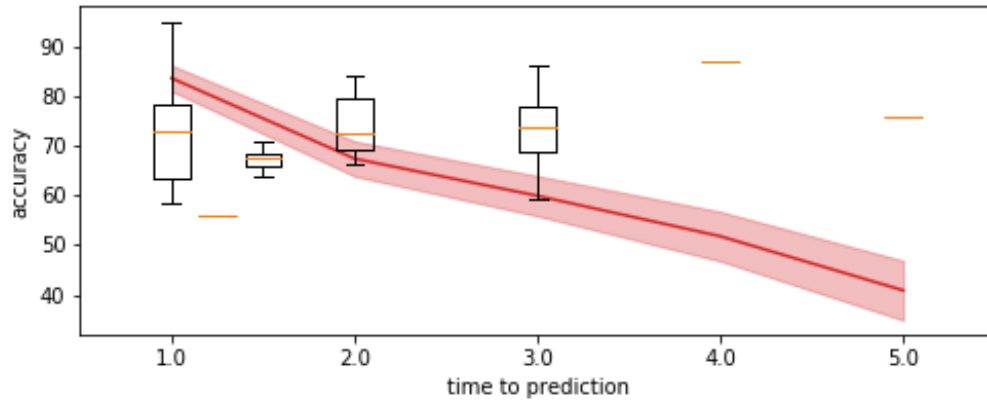


FIGURE 1.4 – Evolution of the performance with respect to the time to prediction. Box plots represent the accuracy of the articles using ADNI. The straight line represents the accuracy of predicting that all individuals remain MCI, corresponding to the proportion of MCI individuals in ADNI staying MCI at the corresponding follow-up time. The shaded area corresponds to the 90% confidence interval of this percentage.

we, as authors, should not only focus on our best performing method, but report also other attempts. Second, it might not be such a problem that innovative methodological works that do not result in a higher performance are published also, provided that the prediction performance is not used to argue about the interest and validity of the method. The machine learning field has the chance to have simple metrics, such as AUC or accuracy, to compare different methods on an objective basis. However, we believe that one should use such metrics wisely not to discourage the publication of innovative methodological works even if it does not yield immediately better prediction performance, and not to overshadow the need to better understand why some methods work better than others.

1.5.1.3 Use of features of test subjects

Feature embedding is performed on the whole data set in 6.8% of experiments, meaning that the features of the test individuals are used for feature embedding during the training phase. As the diagnosis of the test individuals is often not used for feature embedding, as it is for feature selection, performing it on test individual can be considered a less serious issue than for feature selection. It however requires to re-train the algorithm each time the prediction has to be made on a new individual, which is not suited for a use in clinical practice.

1.5.1.4 Use of the diagnosis date

In 5.6% of the experiments, the date of AD diagnosis is used to select the input visit of pMCI individuals, for training and testing. As explained in section 1.2.3, this practice can prevent the generalization of the method to the clinical practice, as the progression date of test individuals is by definition unknown.

These type of experiments answer the question "may one detect some characteristics in the data of a MCI patient 3 years before the diagnosis which, at the same time, is rarely present in stable MCI subjects?". Which should not be confused with : "can such characteristics predict that a MCI patient will progress to AD within the next 3 years". What misses to conclude about the predictive ability is to consider the MCI subjects who have the found characteristics and count the proportion of them who will not develop AD within 3 years.

This confusion typically occurred after the publication of DING et al. (2018). The paper attracted a great attention from general media, including a post on Fox News (WOOLLER, 2018), stating "Artificial intelligence can predict Alzheimer's 6 years earlier than medics". However, the authors state in the paper that "final clinical diagnosis after all follow-up examinations was used as the ground truth label", thus without any control of the follow-up periods that vary across subjects. Therefore, a patient may be considered as a true negative in this study, namely as a true stable MCI subject, whereas this subject may have been followed for less than 6 years. There is no guarantee that this subject is not in fact a false negative for the prediction of diagnosis at 6 years.

1.5.1.5 Choice of time-to-prediction

We find that 22.6% of experiments work on separating pMCI from sMCI, regardless of their time to progression to dementia. We advise against this practice, as the temporal horizon at which the individuals are likely to progress is an important information in clinical practice. Methods predicting the exact progression dates, such as what is asked in the Tadpole challenge (MARINESCU et al., 2018), should be favored over methods predicting the diagnosis at a given date.

The other experiments have set a specific time to prediction, often between 1 and 3 years, meaning that they intend to predict the diagnosis of the individual at the end of this time interval. Figure 1.4 shows the evolution of the accuracy of these methods tested on ADNI with respect to the time to prediction. The time to prediction did not have a significant effect on AUC, accuracy, balanced accuracy, specificity nor sensitivity. Figure 1.4 also shows the accuracy that one would get on ADNI when using a constant prediction, that is predicting that all individuals stay MCI at future time points. The accuracy of this constant prediction has been

computed using the proportion of MCI remaining stable at each visit. We show that most methods predicting the progression to AD within a short-term period smaller of 3 years do not perform better than this constant prediction. We therefore advise to use a time to prediction of at least 3 years, as for a shorter time interval the proportion of MCI individuals progressing to AD is small, predicting that all individuals remain stable therefore gives a better accuracy than most proposed methods.

This fact also shows that the accuracy may be arbitrarily increased by using a cohort with a large proportion of stable subjects. The algorithm may then yield high accuracy by mimicking a constant predictor. This effect may be alleviated by optimizing the balanced accuracy instead of the accuracy.

1.5.1.6 Problem formulation and data set choice

A common theme that arises from the previous issues is that the methods are not always designed to be the most useful in clinical practice. It is for example true of methods that do not use a specific time-to-prediction, or that use the date of AD diagnosis to select the included visits.

More generally, we think the most useful decision support system should not only focus on Alzheimer's disease but perform differential diagnosis. Clinicians do not usually need to distinguish between individuals who will develop AD and individuals who will not develop any neurological disorder. They most likely need help to determine which disorder an MCI individual is likely to develop. Unfortunately, no widely available data set allows the development methods for differential diagnosis to date. Methods focusing on AD should therefore target individuals who have already been identified as at risk of developing AD, by providing insight on the date at which this conversion is likely to happen. Such methods could be trained on MCI subjects that are at risk to develop Alzheimer's disease, defined for instance as the ones who have a MMSE of 27 or smaller and are amyloid positive. In addition to being closer to what can be expected in clinical practice, such data sets of at risk subjects should include a larger proportion of pMCI, leading to a better performance compared to the constant prediction. For example in ADNI, 71.6% of MCI subjects stay stable 2 years after inclusion, whereas this proportion drops to 53.7% for MCI subjects who are amyloid positive and have a MMSE of 27 or lower.

The diagnosis of Alzheimer's disease highly depends on the clinical practice, and varies greatly across sites and countries (BEACH et al., 2012). Therefore, the short-term prediction of progression to Alzheimer's disease are unlikely to generalize well outside the well controlled environment of a research study. An interesting alternative may be to predict the changes in the imaging or clinical biomarkers

in time rather the change in diagnosis, such as proposed by KOVAL, BÔNE et al. (2018) and IDDI et al. (2019).

1.5.2 Proposed guidelines

In order to ensure that the proposed method is useful for clinical practice and that the evaluated performance reflects what could be expected in real life, we propose a list of attention points :

- Separate train and test data sets by using independent cohorts or, if not available, cross-validation.
- No element of the test data set, both labels and features, should be used except for performance evaluation.
- Always pre-register the time window within which one aims to predict conversion to AD, or predict the date of progression.
- Use a large data set or pool different cohorts to obtain a large data set.
- Define a cohort that best reflects the future use of the method in clinical practice. For instance, select subjects that will be considered as at risk of developing the disease rather than all possible ADNI subjects.
- Systematically benchmark the method against the prediction that the subjects will remain stable over time.

1.6 Conclusion

We conducted a systematic and quantitative review on the automatic prediction of the evolution of clinical status of MCI individuals. We reported results from 234 experiments coming from 111 articles. We showed that studies using cognitive variables or FDG PET reported significantly better results than studies that did not. These modalities should be further explored, cognition because it can be easily collected in clinical routine, and FDG PET for the interest it might represent for the imaging community and for increasing our understanding of the disease. On the other hand, we showed that using solely T1 MRI yields a significantly lower performance, despite the great number of methods developed for this imaging modality. These findings call into question the role of imaging, and more particularly of MRI, for the prediction of the progression of MCI individuals to dementia. It would therefore be interesting to shift our focus towards other modalities. More specific cognitive tests could be created, and the impact of using digitized tests, that can be frequently used at home by the patients themselves, should be studied.

We identified several key points that should be checked when creating a method which aims at being used as a clinical decision support. When possible, an

independent test set should be used to evaluate the performance of the method, otherwise a test set can be separated by carefully splitting the cohort. In any case, the test individuals should not be used to make decisions regarding the method, such as the selection of the features or parameter tuning. The time window in which one aims at predicting the progression to AD should be pre-registered, as the temporal horizon at which an individual is likely to progress to AD is a useful information for clinicians. Alzheimer's disease being a very slowly progressive disease, algorithm performance should be systematically compared with the prediction that no change will occur in the future. We have shown indeed that the constant prediction may yield very high performance depending on the time frame of the prediction and the composition of the cohort. Finally, the cohort on which the method is tested should be carefully chosen and defined, so as to reflect the future use in clinical practice as best as possible. At a time where one has great expectation regarding the use of artificial intelligence to support the development of precision medicine, it becomes urgent that the field of AD research adopts state-of-the-art standards and good practices in machine learning.

Chapitre 2

Prediction of future cognitive scores and dementia onset in Mild Cognitive Impairment patients

This chapter is in preparation for submission as a journal article. Results were also published in two conference abstracts :

- **Manon Ansart**, Ninon Burgos, Olivier Colliot, Didier Dormont and Stanley Durrleman. 2019. “Prediction of Future Cognitive Scores and Dementia Onset in Mild Cognitive Impairment Patients.” In **Annual Meeting of the Organization for Human Brain Mapping – OHBM 2019**. <https://hal.inria.fr/hal-02098427>
- **Manon Ansart**, Igor Koval, Anne Bertrand, Didier Dormont, and Stanley Durrleman. 2018. “Design of a Decision Support System for Predicting the Progression of Alzheimer’s Disease.” **Alzheimer’s & Dementia** 14 (7) : P433. <https://doi.org/10.1016/j.jalz.2018.06.371>.

2.1 Introduction

Alzheimer’s disease (AD) is characterized by changes in brain structure and cognition that can be observed before AD diagnosis, in individuals with a Mild Cognitive Impairment (MCI). Some MCI subjects progress to AD (progressing MCI, or pMCI), whereas other individuals are diagnosed with other conditions or keep an MCI clinical status (stable MCI subjects, or sMCI). Identifying MCI individuals who will develop AD is a crucial challenge, as it can impact patient care, and allow the development of new therapeutic strategies targeting the earliest stages of AD.

Several methods, described in Chapter 1 have been proposed to automatically predict the future diagnosis of MCI subjects. A large number of methods focus

on the use of T1 magnetic resonance imaging (T1 MRI), to identify patterns differentiating pMCI from sMCI subjects (BEHESHTI, DEMIREL et MATSUDA, 2017; MINHAS et al., 2018; KAUPPI et al., 2018). CUINGNET et al. (2011) and GÓMEZ-SANCHO, TOHKA et GÓMEZ-VERDEJO (2018) study the effect of the choice of MRI features on the prediction of MCI progression. Other methods focus on the use of neuroimaging more broadly, by integrating features of different modality, in particular ^{18}F 2-fluoro-2-deoxy-D-glucose (FDG) positron emission tomography (PET) (VIVAR et al., 2018; SAMPER-GONZALEZ et al., 2019), and others use cognitive features as well (KAUPPI et al., 2018; MUBEEN et al., 2017; MORADI et al., 2015). These methods rely on the use of machine learning algorithms, which are trained on the input features to directly classify sMCI versus pMCI subjects. They are often difficult to interpret, and we believe this black box effect has prevented the adoption of these methods in clinical practice. We aim at answering this need for interpretability, by proposing a method that first predicts the changes of cognition of the subjects, and uses it to predict the corresponding diagnosis. This method reduces the black box effect by giving an understanding of how the final prediction is made, and provides a more comprehensive view of the future patient state to the clinician.

Alzheimer's disease is a slow progressive disease, with subtle short term changes. Studying the changes visible in a subject over multiple visits can therefore be more informative than looking at one time point only. Taking several visits into account can however be challenging, as different subjects have a different follow-up duration, and they can be observed at different time points. We propose several methods for using more than one past visit for the prediction of MCI progression, and compare the performance they yield with the performance one can obtain by using one past visit only.

We have explained in Chapter 1 that it is important to pre-register the time window in which one aims at predicting the progression to AD. Individuals who are observed for a duration shorter than the defined time window and did not progress to AD should not be included, as they might still be diagnosed with AD in the considered time window, after their last observation. Following this recommendation, we compare our different approaches on a one year prediction. Because we also showed that prediction at more than 2 years is more relevant, we compare prediction we obtain on a one year, two years and three years interval.

Comparing a prediction method with others can be challenging, as different methods are tested on different cohorts, varying in terms of design, number of subjects, prediction time window and proportion of individuals progressing to AD. The Tadpole challenge aims at proposing a common framework for prediction evaluation. The participants were asked to predict the future diagnosis of ADNI3

roll-over participants, with time to prediction ranging from several months to 1.5 years for the first results. We participated in this challenge and report here the performance we obtained with our method, in order to compare it to other proposed methods evaluated in the same settings.

2.2 Materials and methods

2.2.1 Cross-sectional framework

2.2.1.1 Description

We propose a method composed of two steps (Figure 2.1). In a first step, we predict the future of cognitive scores (Alzheimer’s disease assessment scale cognitive sub-scale (ADASCog), Mini-mental state examination (MMSE), Rey auditory verbal learning test (RAVLT) and Clinical dementia rating (CDR)) at time $t + \Delta t$ using MRI extracted volumes (whole brain, entorhinal, fusiform, mid temporal, ventricles and hippocampus volumes), socio-demographic information, APOE genotype and the cognitive scores at time t . This prediction is performed using a Ridge regression (HOERL et KENNARD, 1970), and is trained on the MCI subjects for which 2 visits separated by a Δt interval are available, such that the first visit is associated with a MCI diagnosis. When several pairs of visits are available for one subject, the last pair of visits is used, so as to have as much past visits available as possible for the longitudinal methods. To perform the prediction on test MCI subjects, the last available visit is used.

In a second step, we use the features predicted in the first step to estimate the diagnosis at the same time point ($t + \Delta t$). This prediction is performed using a Gaussian kernel Support Vector Machine (SVM), which is trained on all the available visits of MCI training subjects and AD subjects.

2.2.1.2 Inclusion of additional features

We consider different approaches to improve this prediction pipeline. We first consider using more input features : FDG PET SUVR and detailed MRI features (regional cortical thickness and white matter volumes). As changes in FDG PET and in MRI should appear before changes in cognition, these additional features could provide early markers of the state of the patients and hence improve the prediction of cognition evolution.

We also consider learning the regression on different groups, depending on the age (< 65 years old and > 65 years old), the amyloid status or the APOE genotype.

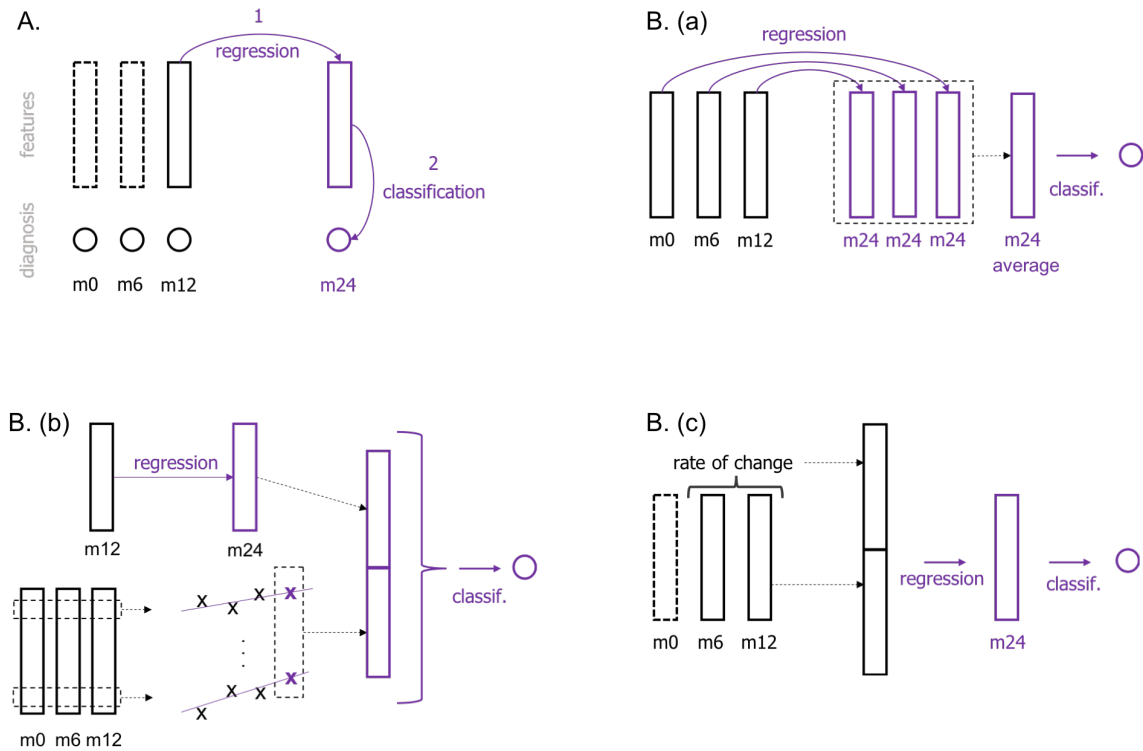


FIGURE 2.1 – Proposed approaches : A. cross-sectional framework ; B. longitudinal framework with 3 propositions : (a) averaging, (b) stacking, (c) rate of change.

Less subjects would therefore be used to train each regression, however the groups are more homogeneous so the regression should fit each individual better.

2.2.2 Longitudinal frameworks

The cross-sectional framework only uses the last available visit for prediction. We consider different strategies in order to incorporate longitudinal information in this pipeline.

2.2.2.1 Averaging approach

In a first approach, we perform the prediction of the future cognitive scores using different time points as input. For each subject, the prediction is performed using each past visit independently, and the predictions are averaged to obtain a global prediction, that is then used to predict the corresponding diagnosis. This approach is described in Figure 2.1 B. (a) and is referred to as the averaging approach. As a regression is trained for a given time interval between the prediction and the input features, a regression is trained for each interval encountered in the test data set. For example, if in the test data set subjects are followed for a maximum of 5 years, 11 regressions will be trained, for time intervals going from 1 to 5

years with a step of 6 months. In an ideal setting, the optimal combination of the predictions made using the each past visit would be automatically performed. It would however imply to use the same number of visits for each subjects. We therefore choose to use a simple average as it allows for the combination of a different number of prediction for each subject, and to hence use all available past visits.

2.2.2.2 Temporal regression and stacking

In a second approach, we perform a time linear regression to predict the next time point of each cognitive score. The next time point is thus predicted using all the previous time points of the given subject, as one linear regression is trained for each subject, hence allowing the inclusion of a different number of past visits in each one. This prediction of the cognitive scores is then combined with the prediction performed in the cross-sectional framework by stacking them in one feature vector, and this vector is then used to predict the corresponding diagnosis. This approach is referred to as the stacking approach, and is described in Figure 2.1 B. (b).

2.2.2.3 Rate of change approach

In a third approach, we compute the rate of change of all the input features between the two last visits. The prediction of the next time point is then performed using the input features and their rate of change, using a ridge regression. The diagnosis prediction is performed as described in the cross-sectional framework. This approach, referred to as the rate of change approach, is described in Figure 2.1 B. (c).

2.2.3 Experimental setup

2.2.3.1 Data set

We evaluate our method on the MCI patients of the ADNI database. We compare the different approaches on a 1 year prediction, using for each subject the latest pair of visits separated by a 1 year interval, so as to have as many past visit available as possible for the longitudinal approaches. Our data set contains 411 subjects with such a pair and a MCI diagnosis before the latest visit. We define sMCI subjects as subjects who remain stable at one year (354 subjects, 86.1%), and pMCI subjects as subjects who progress to AD at 1 year (57 subjects, 13.9%).

In order to increase the number of visits with an associated AD diagnosis, 316 AD subjects from the ADNI study are also included in the training set of the prediction of the diagnosis from the cognitive scores.

We also evaluate the performance of the best approach on a prediction at a 2 year and 3 year interval, for which 354 and 219 MCI subjects are available respectively, among whom 25.4% and 30.1% respectively progressed to AD.

2.2.3.2 Validation procedure

Performance measures are obtained by splitting the cohort 50 times into a training (70%) and test (30%) set. The parameters of the ridge regression and SVM are tuned within a nested 5-fold cross validation. As the cross-validation provides 50 performance measures for each method, the mean performance and its standard deviation are computed. Two-tailed t-tests are used to compare method performance. As a comparison with our method, we predict the diagnosis at time $t+\Delta t$ by using a linear SVM on the features at time t directly.

2.2.4 TADPOLE challenge

The TADPOLE challenge consists in the prediction of future clinical status, ADASCog score and ventricle volume in rollover individuals in the ADNI study. Participants were asked to make monthly predictions from January 2018 to December 2022. The previously described framework was designed to make predictions 1 year after the last visit, and is easily extended to make predictions at a Δt interval. Several of such methods are trained in order to predict the future of the cognitive scores at time points 6 months apart for each subject. Monthly predictions of the cognitive scores are then obtained using linear interpolation, and the monthly values are used as input for the classification. This extension allows to obtain monthly predictions for each subject. The prediction of the ventricle volume was performed using the same method as for cognitive score prediction.

The prediction of the diagnosis was evaluated using the multiclass area under the receiver operating curve (mAUC), defined in HAND et TILL (2001).

2.3 Results

2.3.1 Cross-sectional framework

2.3.1.1 Proposed approach

Results are presented in Table 2.1 and the mean absolute errors (MAE) are shown in Figure 2.2. The proposed cross-sectional approach results in an AUC of 87.9 ± 2.7 , whereas direct classification gives an AUC of 86.6 ± 2.2 . Although the proposed approach performs significantly better ($p < 0.01$) than direct classification, the difference is small. The interest of the method is not only to predict

		Method	AUC	bacc
		Proposed approach	87.9 ± 2.7	77.1 ± 6.1
		Direct classification	86.6 ± 2.2 **	77.4 ± 6.4
Extra features	groups	+ FDG PET	86.7 ± 3.1 *	75.1 ± 3.9
		+ detailed MRI	88.1 ± 2.8	76.9 ± 5.0
		Age	87.7 ± 2.9	74.3 ± 7.2 *
		Amyloid status	87.5 ± 2.9	74.6 ± 5.8 *
		APOE genotype	87.4 ± 3.3	77.2 ± 4.1
Longitudinal		Averaging	83.3 ± 3.9 ***	75.4 ± 5.8
		Stacking	87.6 ± 3.3	75.9 ± 5.7
		Rate of change	87.8 ± 3.1	77.9 ± 6.5

TABLE 2.1 – Area Under the ROC curve (AUC) and balanced accuracy (bacc) obtained on the tested approaches, in the form mean ± standard deviation. Significant differences between the proposed approach and the other method is shown as : *= significant at the 0.05 level; **= significant at the 0.01 level; ***= significant at the 0.001 level

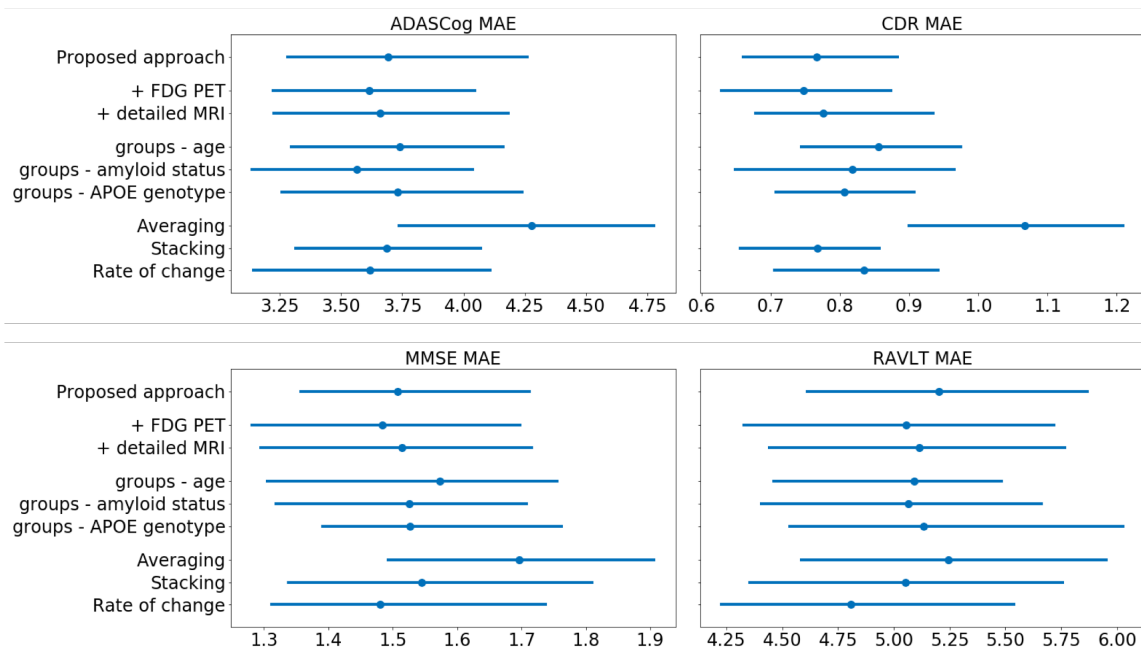


FIGURE 2.2 – Mean absolute error (MAE) of the prediction of the four cognitive scores at one year. The dots correspond to the mean performance and the lines to the 95% confidence interval

the diagnosis but also the value of the cognitive performance in the future. The prediction of the MMSE yields an MAE of 1.51 ± 0.13 (on a scale of 30), and The prediction of ADASCog yields and MAE of 3.69 ± 0.28 (on a scale of 80).

2.3.1.2 Additional features

Additional features are included in an attempt to improve the prediction. As change in FDG-PET and MRI appear before changes in cognition, one could expect that including these features would improve the prediction of the evolution of cognitive scores. However, including FDG-PET in the features lead to a significant decrease in AUC (86.7 ± 3.1 , $p < 0.05$), and including detailed MRI features lead to a non-significant increase in AUC (88.1 ± 2.8 , $p > 0.05$). The significant decrease in AUC induced by the inclusion of FDG-PET might be explained by the reduction of the number of available subjects. As some subjects do not have an FDG-PET value at the used time points, less subjects are including for training the algorithm, which can lower its performance (DOMINGOS, 2012; ANSART, EPELBAUM, GAGLIARDI, COLLIOT, DORMONT, DUBOIS et al., 2019).

2.3.1.3 Building regression groups

We build different regression groups, according to age, amyloid status or APOE genotype. As subjects within these groups are more similar than in the whole cohort, one may expect that they exhibit more similar progression patterns. In this case, the regression would better fit each individual and the cognitive scores would be better predicted. Building different regression groups however leads to a non-significant decrease in AUC for all groups. This decrease in AUC may be explained by the reduction of the number of subjects available to train each regression. The decrease in performance due to a smaller data set seems to be greater than the possible increase due to more homogeneous populations.

2.3.2 Longitudinal frameworks

2.3.2.1 Averaging approach

The averaging longitudinal approach uses all the past visit of each subjects by averaging the predictions from the different visits. Although this approach takes advantage of all the information available for each subjects, it leads to a significantly lower AUC than the cross-sectional approach (83.3 ± 3.9 (std), $p < 10^{-9}$). Long-term predictions are less precise than short-term ones (see subsection 2.3.3). So the predictions from the earliest time-points may not add relevant information, hence leading to a lower performance than the one obtained using the latest visit

only. This effect might be reduced by affecting a different weight to the various time points.

2.3.2.2 Stacking approach

	Cross-sectional prediction coefficient	Time regression coefficient
MMSE	0.93 (0.13)	0.09 (0.09)
ADASCog	0.80 (0.13)	0.19 (0.09)
CDR	0.63 (0.11)	0.42 (0.13)
RAVLT	0.89 (0.14)	0.09 (0.09)

TABLE 2.2 – Coefficients of the regression combining the cross-sectional prediction and the time linear regression in the stacking approach. Data are mean (standard deviation).

The stacking approach combines the cross-sectional prediction of the cognitive scores with a longitudinal one, performed using a time linear regression for each subject. The two predictions are included in one feature vector which is used then as input for the classification. As opposed to the averaging vector, the weight given to longitudinal prediction and to the cross-sectional one is thus automatically learned during the classification. This methods lead to an AUC of 87.6 ± 3.3 , which is not significantly different from the cross-sectional approach ($p > 0.5$). As the prediction of the cognitive score is better using the cross-sectional method than using a time linear regression for each subject, the weight given to this second feature set is low, as shown in Table 2.2, and the prediction using this method is close to the prediction obtained using the cross-sectional method solely.

2.3.2.3 Rate of change approach

The rate of change approach combines the input features and their rate of change in a larger feature vector, used as input to predict the evolution of the cognitive scores. This longitudinal approach is the simplest one. It combines only two time points, but using all the input features of all the subjects, as opposed to the stacking approach in which the longitudinal prediction of the cognitive scores is performed using only the past values of this cognitive score for the given subject. This approach yields an AUC of 87.8 ± 3.1 , which is not significantly different from the AUC of the cross-sectional framework ($p > 0.05$). As this simple approach

does not result in any increase in performance, we can suppose that including longitudinal information does not lead to a better prediction of the future diagnosis than using one past visit only. Similar conclusions can be drawn using the MAE.

2.3.3 Prediction at different temporal horizons

	1 year	2 years	3 years
Number of individuals	480	354	219
% of MCIc	13.9	25.4	30.1
AUC	87.9 ± 2.7	87.9 ± 2.8	88.8 ± 4.3
bacc	77.1 ± 6.1	78.9 ± 4.4	78.0 ± 6.9
MMSE MAE	1.51 ± 0.13	1.87 ± 0.13	2.20 ± 0.18
ADAS MAE	3.69 ± 0.28	4.51 ± 0.34	5.38 ± 0.52
CDR MAE	5.20 ± 0.38	5.76 ± 0.36	6.16 ± 0.49
RAVLT MAE	0.77 ± 0.07	1.11 ± 0.09	1.42 ± 0.14

TABLE 2.3 – Performance of the cross-sectional approach for prediction at 1 year, 2 years, 3 years, in terms of Mean Absolute Error (MAE), Area Under the ROC curve (AUC) and balanced accuracy (bacc).

As no other method outperforms the cross-sectional approach on the one year prediction, this approach was used for prediction at other time intervals. Performance of prediction at one year, 2 years and 3 years are presented in Table 2.3. Compared to the 1-year prediction, the MAE of the prediction of the cognitive score at 2 years and 3 years is significantly higher for all cognitive scores ($p < 10^{-10}$). This result is not surprising, as predicting the change in cognition further in time is more difficult. However, the AUC of the 2-year prediction is equal to the AUC at one year, and the AUC at 3 years is better than at 1 year, although not significantly ($p > 0.1$). This shows that even though predicting the cognitive scores further in time is more difficult, the performance of the final prediction, based on the predicted cognitive scores, is more robust. There is a range of of the cognitive scores that is associated with the same diagnosis.

In order to compare our method with other approaches tested on the ADNI data set, we use the review of automatic method for predicting the progression of MCI presented in Chapter 1. We extracted the methods tested on the ADNI

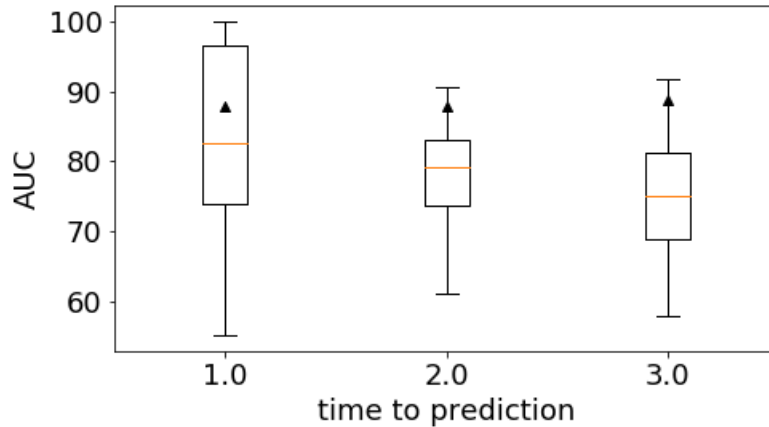


FIGURE 2.3 – Comparison of our performance with the AUC of methods identified in the literature review for different times to prediction. Box plots correspond the performance of other methods, the whiskers going from the minimum to the maximum value. Triangles represent the mean AUC of our approach. AUC : Area Under the ROC Curve.

data set with a corresponding AUC, and using a time to prediction of 1 year (18 methods), 2 years (24 methods) or 3 years (29 methods), corresponding to a total of 71 experiments extracted from 34 articles. We compare their performance with the one we obtain of these 3 prediction intervals in Figure 2.3, where the AUC of other proposed methods is shown as box plots, and the AUC of our method as triangles. Figure 2.3 shows that, while our method is just above the median of 1-year prediction AUC, it performs well compared to other method on the 2-year and 3-year predictions. As shown in Chapter 1 and in Table 2.3, few subjects convert at 1 year, therefore the accuracy of methods predicting diagnosis at 1 year is often below the accuracy one can obtain by assuming all subjects stay MCI. Performing a prediction at a 2 year or 3 year horizon is therefore more suited, and on these time intervals our method performs well compared to other methods.

2.3.4 TADPOLE challenge

We evaluated our method by submitting to the TADPOLE challenge. The past visits of 380 cognitively normal (CN) individuals, 422 MCI individuals and 475 AD individuals was available. The last observation of the individuals was made between 2007 and 2017, and predictions had to be made between 2018 and 2020, resulting in times to prediction of 6 months to 2.5 years for the subjects observed most recently. The clinical diagnosis of AD and CN subjects is not expected to change, especially in such a short period of time, the performance therefore relies mostly on MCI prediction. Because of the inclusion of CN and AD subjects in this

challenge, and because several predictions are made for each subject, the performance cannot be easily compared with the one obtained on the prediction of MCI progression.

Results obtained on the TADPOLE challenge are presented in Table 2.4. Our approach resulted in an mAUC of 90.2% and a balanced accuracy of 82.7% for diagnosis prediction, and in a MAE of 5.57 for ADASCog prediction and of 0.52 for ventricle volume prediction. The winners of the competition achieved an MAUC of 93.1%. Our overall rank was of 6 out of 52 participants, and first as a university team.

	mAUC	bacc	ADASCog MAE	Ventricles MAE
Proposition	90.2 %	82.7 %	5.57	0.52

TABLE 2.4 – Results obtained on the TADPOLE challenge. mAUC : multiclass area under the receiver operating curve; bacc : balanced accuracy; MAE : mean absolute error.

2.4 Discussion

2.4.1 Cross-sectional experiments

Our results show that the number of individuals available for training is a key factor influencing the performance. We can indeed hypothesize that this effect leads to a lower performance when using FDG PET as it reduces the number of individuals that can be used for training. It may also explain why forming more homogeneous groups does not lead to an increase in performance. A decrease in performance due to a decrease in data set size has already been reported in DOMINGOS (2012), as well as in our work on amyloidosis prediction (ANSART, EPELBAUM, GAGLIARDI, COLLIOT, DORMONT, DUBOIS et al., 2019) and in our review of automatic methods for predicting the evolution of MCI, presented in Chapter 1.

Including detailed MRI features did not lead to a significant increase in AUC. We showed in the in Chapter 1 that including MRI features does not lead to a significant increase in prediction of the progression of MCI patients. Simple but relevant MRI features were already included in the baseline model (whole brain, entorhinal, fusiform, midtemporal, ventricles and hippocampus volumes). Our results show that more detailed features do not bring information that is useful to predict MCI progression and that is not already available in the reduced set. Exploring

methods performing automatic feature extraction on images, in end-to-end learning approaches or using auto-encoders, could allow to better identify relevant information in the images themselves.

2.4.2 Longitudinal frameworks

None of the longitudinal frameworks lead to a performance significantly better than the one obtained using longitudinal data. More sophisticated longitudinal methods, such as proposed in KOVAL, ALLASSONNIÈRE et DURRLEMAN (2019), could lead to a better use of longitudinal data. However, the fact that even the inclusion of the rate of change in the inputs does not lead to an increase in AUC tends to show that one time-point information is enough for this prediction. This results meet conclusions made in AKSMAN (2017) for the classification of AD and in SCHUSTER et al. (2015) for progressive diseases in general. Although this finding can be disappointing from a methodological view point, from a clinical view point it is rather positive. Methods based on longitudinal data require patients to be followed for a certain period of time before a prediction can be made. In clinical practice however, a immediate prediction, made as soon as the patient arrives, is more valuable.

2.4.3 Interpretability

The main interest of the method we propose lies in the increase in interpretability : as the diagnosis prediction is made from the prediction of cognitive scores, it is easier for the clinician to understand it, and have an overview of the future patient state. As for longitudinal approaches, inclusion of additional features and building of different regression groups did not lead to a significant increase in AUC, our final method is simple, hence easy to use in clinical practice, and more understandable for clinicians. Despite its simplicity, our method performed well on the TADPOLE challenge, with a rank of 6 out of 52, and an MAUC close to the winning one. It also performs well on the prediction at 2 and 3 years compared to other proposed methods in the literature.

2.4.4 TADPOLE challenge

Comparison of prediction methods for the evolution of MCI can be a difficult task, as such methods are often evaluated in different settings, on different data sets, and with different goals. The main interest of a challenge such as TADPOLE is to propose a common framework in which different prediction methods can be compared. Data leakage and over-fitting of the test set is often not possible : in the

TADPOLE challenge, participants had to predict data at future time points before they were acquired.

An issue that should be mentioned when considering such challenges is their transposition to the clinical practice. Performance is often the only factor considered for participant ranking. Interpretability for clinicians, the cost and availability of the input features are not taken into account, whereas they can greatly impact the usability of the method in practice. Performance alone is not enough, and usability should be kept in mind when building prediction methods and challenges. An effort has been made in this regard in the TADPOLE challenge. The participants had to predict the evolution of the ADASCog and ventricle volumes, which can bring a better insight on the patient future evolution than a diagnosis category. A separate ranking was also done for cross-sectional methods which are easier to use in clinical routine. Finally, as opposed to previous challenges, participants had to make monthly predictions, hence estimating a date of progression for MCI subjects who are expected to progress to AD. This type of information is crucial for clinicians, and is often left out of automatic prediction methods that separate sMCI for pMCI at no specific temporal horizon.

The definition of the test data set can however be questioned. AD and CN subjects were included, when their diagnosis is not expected to change much in the short term. At the time of the first TADPOLE results, the time to prediction of the subjects who were the most recently observed was of 6 months to 2 years. As shown in section 2.3.3 and in Chapter 1, on such short term prediction the proportion of MCI subjects remaining stable is quite high, predicting that all MCI subjects remain stable can therefore lead to a good performance. This definition of the test data set therefore favors conservative methods, predicting few changes in diagnosis.

Interpretability and usability are not always taken into account when developing automatic methods to predict MCI evolution. Despite the efforts made in the TADPOLE challenge, or in methods such as ANTILA et al. (2013), the usage of machine learning algorithms sometime leads to seeking higher and higher performances, at the cost of a loss in usability. As the ultimate goal of such methods is to be integrated in clinical routine in order to improve patient care, future research would gain at ensuring the usability of automatic methods in clinical practice.

2.5 Conclusion

We proposed a method for predicting the future diagnosis of MCI subject by first predicting their change in cognitive scores. This two-step prediction reduces

the black-box effect of machine learning methods and is easier to interpret for clinicians. It also provides a more complete view of future patient characteristics. The prediction of the future cognitive scores can be used to perform patient clustering and to extract a sub-group of patients with similar characteristics. It can also be used to tailor patient care in a personalized approach. We evaluated our method on 1, 2 and 3 year prediction, showing that predicting cognitive scores on the long term is more difficult, but that diagnosis prediction stays robust. We also evaluated our method on the Tadpole challenge, resulting in a competitive performance.

We compared several methodological options in this prediction framework. We showed that using detailed MRI features did not improve the performance compared to using simple ones, neither did including FDG PET features. Training the regression on more homogeneous patients groups, created based on age, Apoe4 and amyloid status, did not increase the performance either. Finally, performing the prediction based on several past visits for each individual did not improve the performance compared to using one past visit only. Overall, we showed that using more complex features, which can be less accessible in clinical practice, did not lead to a better prediction than using the more simple settings.

Chapitre 3

Reduction of Recruitment Costs in Preclinical AD Trials : Validation of Automatic Pre-Screening Algorithm for Brain Amyloidosis

This chapter has been published as a journal article in *Statistical Methods in Medical Research* :

- **Manon Ansart**, Stéphane Epelbaum, Geoffroy Gagliardi, Olivier Colliot, Didier Dormont, Bruno Dubois, Harald Hampel, Stanley Durrleman, and for the Alzheimer’s Disease Neuroimaging Initiative and the INSIGHT-preAD study. 2019. “Reduction of Recruitment Costs in Preclinical AD Trials : Validation of Automatic Pre-Screening Algorithm for Brain Amyloidosis.” **Statistical Methods in Medical Research**, January, 096228021882303. <https://doi.org/10.1177/0962280218823036>.

This work also lead to the following publications :

- **Manon Ansart**, Stéphane Epelbaum, Olivier Colliot, Didier Dormont, Bruno Dubois, Harald Hampel, and Stanley Durrleman. 2017. “Selection of Amyloid Positive Pre-Symptomatic Subjects Using Automatic Analysis of Neuropsychological and MRI Data for Cost Effective Inclusion Procedures in Clinical Trials.” In **The Journal of Prevention of Alzheimer’s Disease**, 4 :P295. <https://hal.inria.fr/hal-02098434v2>.
- **Manon Ansart**, Stéphane Epelbaum, Geoffroy Gagliardi, Olivier Colliot, Didier Dormont, Bruno Dubois, Harald Hampel, and Stanley Durrleman. 2017. “Prediction of Amyloidosis from Neuropsychological and MRI Data for Cost Effective Inclusion of Pre-Symptomatic Subjects in Clinical Trials.” In **Deep Learning in Medical Image Analysis and Multimodal Learning**

for Clinical Decision Support, edited by M. Jorge Cardoso, Tal Arbel, Gustavo Carneiro, Tanveer Syeda-Mahmood, João Manuel R.S. Tavares, Mehdi Moradi, Andrew Bradley, et al., 10553 :357–64. Cham : Springer International Publishing. https://doi.org/10.1007/978-3-319-67558-9_41.

3.1 Abstract

We propose a method for recruiting asymptomatic Amyloid positive individuals in clinical trials, using a two-step process. We first select during a pre-screening phase a subset of individuals which are more likely to be amyloid positive based on the automatic analysis of data acquired during routine clinical practice, before doing a confirmatory PET-scan to these selected individuals only. This method leads to an increased number of recruitments and to a reduced number of PET-scans, resulting in a decrease in overall recruitment costs. We validate our method on 3 different cohorts, and consider 5 different classification algorithms for the pre-screening phase. We show that the best results are obtained using solely cognitive, genetic and socio-demographic features, as the slight increased performance when using MRI or longitudinal data is balanced by the cost increase they induce. We show that the proposed method generalizes well when tested on an independent cohort, and that the characteristics of the selected set of individuals are identical to the characteristics of a population selected in a standard way. The proposed approach shows how Machine Learning can be used effectively in practice to optimize recruitment costs in clinical trials.

3.2 Introduction

3.2.1 Background

Amyloid plaques, together with neurofibrillary tangles, are one of the earliest signs of Alzheimer’s disease (AD), appearing before any cognitive impairment and change in brain structure (DUBOIS, HAMPEL et al., 2016; JACK, KNOPMAN et al., 2010). They are thought to play an important role in the disease, by triggering a cascade of events leading to neuronal loss and cognitive impairment (J. A. HARDY et HIGGINS, 1992; J. HARDY et SELKOE, 2002; J. HARDY et ALLSOP, 1991). This Amyloid cascade hypothesis has been very influential in therapeutic research, as it is hoped that stopping the formation of the plaques will stop the cascade and hence the progression of the disease. Several molecules have been designed

to target these plaques, by preventing the formation of the $A\beta$ peptides, by clearing them or by stopping them from aggregating to form Amyloid plaques (KARRAN, MERCKEN et STROOPER, 2011). Several of these drugs, such as solanezumab (DOODY et al., 2014) and bapineuzumab (SALLOWAY et al., 2014), have been tested on individuals with dementia or with mild cognitive impairments, but did not result in a decrease of the cognitive decline. The focus of clinical trials is therefore now shifting towards pre-clinical and prodromal individuals, as in the A4 study (trial identifier : NCT02008357) and the clinical trial for CNP520 (identifier : NCT03131453). The Amyloid cascade is thought to be a long, progressive process. Slowing down the formation of Amyloid plaques at the beginning of the process, when individuals are not yet cognitively impaired, should have effects on the long run (DOODY et al., 2014; BECKER et GREIG, 2014), whereas on symptomatic individuals cognitive damage has already occurred and might not be reversed.

Setting up clinical trials targeting asymptomatic individuals with amyloid plaques can however lead to important recruitment costs than can be prohibitive, as it is necessary to ensure that all enrolled individuals have amyloidosis (O'BRIEN et HERHOLZ, 2015; WATSON et al., 2014). The presence of amyloid plaques on the brain can be measured using Positron emission tomography (PET), or by measuring the concentration of $A\beta$ protein in the cerebral spinal fluid (CSF). PET scans are very costly (around 1 000€ in Europe, and 5 000\$ in the United-States) and require the injection of a radioactive compound, and CSF measurements require a lumbar puncture, which is an invasive procedure that cannot be considered for systematic screening. When recruiting amyloid positive ($A\beta+$) individuals in a cohort of individuals with dementia, doing a PET scan to every possible individual can be a reasonable solution, as 90% are expected to be $A\beta+$ (CHÉTELAT et al., 2013). However, in an elderly asymptomatic population, only one third of the individuals are $A\beta+$ (CHÉTELAT et al., 2013). This implies that in order to recruit a given number of $A\beta+$ individuals, three times as many individuals should be tested for amyloid positivity. Therefore, doing a PET scan to every recruited individual does not seem to represent a feasible solution for the large-scale recruitment of asymptomatic amyloid positive individuals (WITTE et al., 2015).

We propose a method for recruiting asymptomatic $A\beta+$ individuals for clinical trials, which is composed of two steps, as presented in Figure 1. In a pre-screening phase, we first identify a subpopulation with a higher prevalence of $A\beta+$ individuals than in the original cohort, before doing a PET scan to this sub-population only in a second phase. In order to identify individuals with a higher risk of being $A\beta+$, we propose to use a classifier that has been optimized to minimize the recruitment cost.

3.2.2 Related works

Several methods have been proposed to automatically predict the amyloid status of Cognitively Normal (CN) individuals based on cognitive and socio-demographic information. Mielke et al (MIELKE et al., 2012) use a logistic regression with a default threshold value, and evaluate their method by training and testing the algorithm on the same individuals. Insel et al (INSEL et al., 2016) use a Random Forest and optimize the threshold by maximizing the Positive Predictive Value (PPV) of the algorithm. Maximizing this value implies having a very high threshold value, hence being very selective and increasing the number of false negatives. A very large number of individuals then has to be recruited as input, as many positive individuals are discarded.

Other methods focus on MRI features, such as Tosun et al (TOSUN, JOSHI et WEINER, 2013) who predict amyloidosis in subjects with a Mild Cognitive Impairment (MCI) using an advanced anatomical shape variation measure. Apostolova et al (APOSTOLOVA et al., 2015) also include MRI features by using hippocampus volume and cognitive, ApoE4 and peripheral blood protein information on MCI subjects using an SVM. Ten Kate et al (KATE et al., 2018) use an SVM and tree-based feature selection to predict amyloidosis in CN and MCI subjects using cognitive, socio-demographic, ApoE4 and MRI features. In this paper, we propose to take a cost-effective approach of the amyloidosis prediction, by comparing different methods in terms of cost reduction.

Another approach for reducing clinical trial costs consists in adapting clinical trial design using previous results. Several studies propose to assess treatment efficacy in a retrospective manner, using drug trial cohorts to identify a subgroup of patients responding to treatment (FOSTER, TAYLOR et RUBERG, 2011; QIAN et MURPHY, 2011; Y. ZHAO et al., 2012). On the other hand, other studies propose to do so in a prospective manner, adapting the clinical trial as it is ongoing, by using more advanced methods such as active learning (MINSKER, Y.-Q. ZHAO et CHENG, 2016; SATLIN et al., 2016).

3.2.3 Contributions

Selecting amyloid positive subjects for cohort recruitment requires to find a balance between being very selective, hence discarding a large number of positive individuals on one hand, or being too permissive and doing unnecessary PET scans on the other hand. We propose to take this trade-off into account by optimizing the algorithm for the recruitment cost, which includes both the cost of recruiting a number R of individuals and the cost of doing a confirmatory PET scan to a number S of selected individuals. As R depends on the number of False Negative and

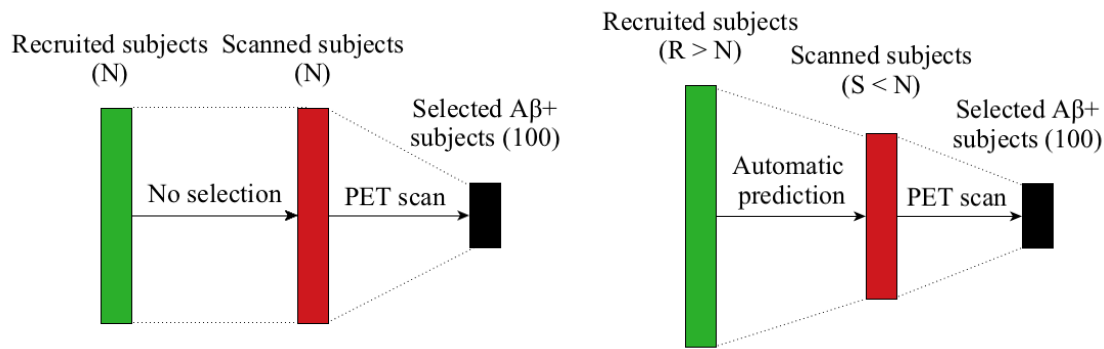


FIGURE 3.1 – Selecting $A\beta+$ individuals : current (left) and proposed (right) process

S on the number of False Positive, both of these measures are taken into account when the cost is minimized.

In this study, we extend and evaluate more in depth the approach we proposed in 2017 (ANSART, EPELBAUM, GAGLIARDI, COLLIOT, DORMONT, HAMPEL et al., 2017). We will compare the performance obtained using different features sets, containing cognitive and imaging features at baseline or over a longitudinal follow-up, and compare performance for a variety of classification algorithms. All the algorithms will be cross-validated to maximize the area under the receiver operating characteristic (ROC) curve (AUC), and the threshold will be chosen to minimize the cost. We will validate our method on three different data sets, corresponding to different disease stages (pre-clinical or prodromal) or recruiting procedures. The performance will be assessed using two different validation procedures : by using cross-validation on each cohort; and by training the algorithm on a first cohort and testing it on a different one. We will then verify that the cohorts created with our method are unbiased, and can be used as inputs for clinical trials.

3.3 Materials and Methods

3.3.1 Cohorts

We are interested in studying the performance of our method on different groups of individuals. To do so, we test the method on three cohorts, noted ADNI-MCI, ADNI-CN and INSIGHT.

The ADNI-MCI cohort contains MCI subjects from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) study. It is an ongoing, longitudinal, multicenter American study carried out in North America, which provides biomarkers, imaging, cognitive and genetic data, for the early detection of AD. It started in 2004

with ADNI1, and two more phases are now available : ADNIGO and ADNI2. A diagnosis is given at each visit, among CN (Cognitively Normal), MCI or AD. MCI subjects have a Subjective Memory Concern (SMC) and an objective memory loss measured by education adjusted scores on Wechsler Memory Scale Logical Memory II, but don't have any impairment in the other cognitive domains, especially in activities of daily living. We only consider visits that have an associated $A\beta$ level, measured with the AV45 PET SUVR (Standardized Uptake Value Ratio) when available, or with the CSF biomarker when no PET scan was performed. individuals that changed Amyloid status during the study are removed. We use the first available visit for each individuals, and a visit at a 12 months interval when studying the impact of longitudinal data. 596 individuals are available in this cohort, among which 62.9% are $A\beta+$.

The ADNI-CN cohort contains CN subjects from the ADNI study. These individuals are cognitively normal, they show no sign of dementia or of cognitive impairment, but they can have a SMC. individuals and visits are selected and $A\beta$ values are taken as in the ADNI-MCI cohorts. 431 individuals are available, among which 37.6% are $A\beta+$.

The INSIGHT cohort contains individuals from the INSIGHT-preAD study. It is an ongoing, longitudinal, mono-centric French study carried out in Paris, France, which aims at studying changes appearing in healthy individuals, over 70 years of age in order to study the very early phases of AD. 318 CN individuals, with normal cognition and memory but who have a SMC, are followed. Cognitive, imaging and genetic data is available for every annual visit. The AV45 PET SUVR is available for every individual and used as the $A\beta$ value. At the time of the analysis, only the first visit is available for each individual. 27.7% of the 318 individuals are $A\beta+$ (n=88).

3.3.2 Input Features

Different sets of features are compared. For all experiments, socio-demographic features (age, gender, education) and ApoE4 are used.

As cognitive assessments are different in ADNI and the INSIGHT-preAD study, different cognitive features are used. For the two ADNI cohorts, the Alzheimer's Disease Assessment Scale - cognitive subscale (ADASCog) is used. The 13 items are aggregated into 4 categories : memory, language, concentration and praxis. For the INSIGHT cohort, the 112 available features, coming from SMC questionnaires and cognitive tests are used. They target executive functions, behavior and overall cognitive skills.

MRI extracted features are also used in order to evaluate their predictive power. The cortical thicknesses are extracted using FreeSurfer for both ADNI and

INSIGHT subjects. The average thicknesses of 72 cortical regions are used, and divided by the total cortical thickness in order to get comparable measures across individuals. The hippocampus volume is extracted using FreeSurfer for the ADNI cohorts, and using SACHA (CHUPIN et al., 2009), an in-house hippocampus segmentation software, for the INSIGHT-preAD study.

The amyloidosis is measured using a PET scan when available and CSF measurements elsewhere. The PET SUVR given by the ADNI and INSIGHT-preAD studies are extracted using different methods. A individual is considered $A\beta+$ when PET SUVR is above 1.1 (CLARK et al., 2012) for ADNI and 0.79 for the INSIGHT-preAD study, or when the concentration of $A\beta$ in the CSF is below 192 pg/ml (SHAW et al., 2009).

3.3.3 Algorithms

Different classification algorithms are used to make the prediction and their performances are compared for the different cohorts, in order to identify an algorithm that would outperform the others. The hyperparameters of all the algorithms are tuned using a cross-validation.

5 algorithms are compared : (1) A Random Forest (BREIMAN, 2001), with validation of the number and the depth of the trees, (2) A logistic regression (J. FRIEDMAN, HASTIE et TIBSHIRANI, 2010), with validation of the threshold, (3) a linear Support Vector Machine (MULLER et al., 2001) (SVM), with validation of the penalty parameter, (4) an adaptive logistic regression (J. FRIEDMAN, HASTIE et TIBSHIRANI, 2000) (AdaLogReg), with validation of the learning rate and the number and depth of the learners, (5) an adaptive boosting (J. H. FRIEDMAN, 2001) (AdaBoost), with validation of the same hyperparameters as for AdaLogReg.

The performance of the algorithms is evaluated using repeated random subsampling validation : the data is repeatedly (50 times) separated into a training set (drawn without replacement) and a test set (corresponding to the data points not used in the training set). We use 70% of the data for training and 30% for testing. For each split, the algorithms are first tuned using a 5-fold validation on the training set to maximize the AUC, then trained on the whole training set with the selected hyperparameters, and applied on the test set in order to get a performance measure. 50 performance measures are therefore obtained, and are used to get a mean performance and a standard deviation. The whole procedure is described in pseudocode in the Supplementary Materials (Algorithm 1).

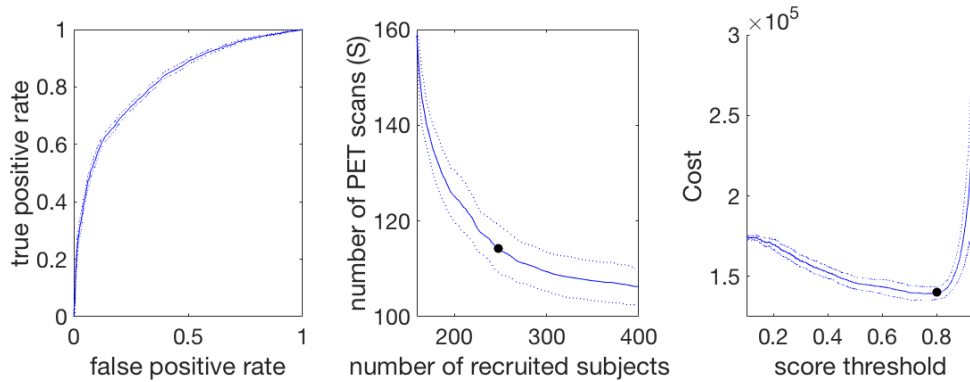


FIGURE 3.2 – Example of ROC curve (left), S vs R curve (middle) and corresponding cost curve (right). The solid curve represents the mean performance, the dotted ones represent the standard deviation and the 2 black dots the points of minimal cost

3.3.4 Performance Measures

Different performance measures are used in order to evaluate different aspects of the methods.

The area under the Receiver Operating Characteristic (ROC) curve (AUC) is used to evaluate the performance of the prediction method. It is used to compare different algorithms, to tune them, and to evaluate the predictive power of different feature sets.

The minimal cost of recruiting 100 individuals is used to measure the practical effect of the method, and to find a balance between the number of recruited individuals and the number of PET scans. In order to compute this minimal cost, the ROC curve is built by changing the algorithm threshold (Fig 3.2, left). For each point on the ROC curve, the corresponding number of individuals to be recruited (R) and the number PET scans (S) is computed (Fig 3.2, middle) as such :

$$S = 100 * \frac{TP + FP}{TP} \quad (3.1) \quad R = 100 * \frac{N}{TP} \quad (3.2) \text{ where}$$

TP stands for number of True Positive, FP for number of False Positive and N is the total number of predictions that have been made. As the true positive rate (TPR) and false positive rate (FPR) depend on the number of True Positive and False Positive which are used to compute S and R, there is a direct match between each point of the ROC curve and the R vs S curve. Consequently, as for the FPR and TPR, R and S should be minimized together and a trade-off has to be made, which is reflected in the total cost.

For each value of S and R, the corresponding cost can be computed, by making the hypothesis that recruiting a individual and getting genetic information and cognitive assessments costs 100€, doing an MRI 400€ and doing a PET scan 1000€. When the cost curve (Fig 3.2, right) is built, the minimum is taken to get the

Data set	Random Forest	Logistic regression	SVM	AdaLogReg	AdaBoost
INSIGHT	67.5 (5.5)	62.7 (6.1) *	62.0 (5.8) *	67.5 (5.7)	67.2 (6.9)
ADNI-CN	69.1 (4.0)	69.5 (4.1)	67.3 (5.0)	66.4 (4.6) *	66.5 (5.1)
ADNI-MCI	82.4 (2.8)	81.9 (2.6)	81.8 (2.7)	80.9 (2.8)	80.5 (3.3) *

TABLE 3.1 – Benchmark of algorithms. * = statistically significantly different from the Random Forest at the 0.05 level after Bonferroni correction for multiple comparisons. Data are : average Area Under the ROC curve (standard deviation). SVM = support vector machine; AdaLogReg = adaptive logistic regression; AdaBoost = adaptive boosting; ROC = receiver operating characteristic.

minimal cost of recruiting 100 individuals, and the corresponding optimal values of S and R are hence known.

It is to be noted that the cost of recruiting 100 individuals in a cohort will depend on the proportion of amyloid positive individuals in the cohort, as the more positive individuals there are, the easier it is. This performance measure is hence useful to evaluate and compare the performance of different methods on one cohort, but it cannot be used to compare the performance of a method across different cohorts.

3.3.4.0.1 Statistical testing Each experiment is performed 50 times with 50 train/-test split, and 50 performance measures are obtained. When we compare two experiments, a two-tailed t-test is performed using the 50 performance measures of each experiment. A p-value is obtained, enabling us to test if the performance of the two experiments is significantly different at the 0.05 level.

3.4 Results

3.4.1 Algorithm and feature choice

3.4.1.1 Algorithm choice

In order to choose the algorithm most suited for this problem, different classification algorithms are tested on the three data sets. Their performance, measured using the AUC, is reported in Table 3.1. These results show that there is no algorithm that outperforms all the others for all cohorts. It is however necessary to make a choice and use the same algorithm on all cohorts. The Random Forest is, for

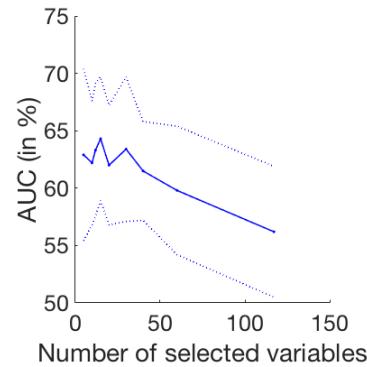


FIGURE 3.3 – Performance variations depending on the number of kept LASSO variables : mean Area Under the ROC curve (plain line) and standard deviation (dashed lines)

all data sets, among the best performing algorithms. It outperforms each other algorithm in one cohort : the Logistic Regression in INSIGHT ($p = 0.001$), the SVM in INSIGHT ($p=0.0001$), the adaptive logistic regression in ADNI-CN ($p=0.03$) and AdaBoost in ADNI-MCI ($p=0.045$). No algorithm significantly outperforms it on any cohort. The Random Forest therefore represent the best algorithm for this classification task.

3.4.1.2 Feature selection for cognitive variables

In the INSIGHT cohort 112 cognitive features are available. Using all of them results in an AUC of 56.2% (± 7.5), which is significantly lower than the performance obtained on the other cohorts because of a less favorable ratio between number of features and individuals, as only 318 individuals are available. We therefore compare different dimension reduction and feature selection methods in order to solve this issue and improve the performance on this cohort.

3.4.1.2.1 Automatic methods Principal Component Analysis (PCA) and Independent Component Analysis (ICA) using fastICA (HYVARINEN, 1999) are first considered, but both lead to an AUC under 52%, whatever the number of selected dimensions.

LASSO feature selection is also considered. In the LASSO, a regularized regression using a l^1 penalty is used, setting some of the feature weights to 0, hence keeping only the most relevant features. A linear regression using LASSO is performed between the input features and the amyloid status in order to select from 5 up to 60 features. The selected features are then used to perform the classification, using a Random Forest. The evolution of the AUC with the number of selected features is presented in Fig. 3.3, showing that the best results are obtained using 15 features. Using the LASSO features selection leads to an AUC of 64.3%

(± 5.2), which is significantly better than the performance obtained using all features ($p < 0.0001$).

3.4.1.2.2 Using expert knowledge In a last analysis, manual feature engineering is considered. Aggregates are formed for each cognitive test, using expert knowledge regarding the tests and the features which are most relevant for AD diagnosis. 26 aggregates are hence built. Using them as input in place of the 112 original cognitive features leads to an AUC of 67.5% (± 5.5), which is significantly better ($p < 0.005$) than the performance obtained using automatic dimension reduction.

3.4.1.3 Use of MRI

We want to assess the prediction power of MRI-extracted features (cortical thicknesses and hippocampus volume) and compare it with the performance obtained using cognitive features. In all experiments, ApoE4 genotype and socio-demographic features are also used as inputs.

We first compared the performance obtained by using only cognitive features on one hand, and only MRI features on the other. As the number of MRI features is large regarding the number of subject, a LASSO feature selection is performed to select 12 variables. The results are presented on lines 1 and 2 of Table 3.2. Using MRI features instead of cognitive scores leads to a significant decrease in the AUC for all cohorts ($p < 0.001$). These results show that the used cognitive features are a better predictor of amyloidosis than the chosen set of MRI features.

Although they are less predictive than cognitive scores, using the MRI features as input along with cognitive scores could lead to better performance. We therefore train the algorithm using both MRI and cognitive features and compare its performance with the ones obtained using solely cognitive scores. The results, presented in line 1 and 3 of Table 3.2, show that including MRI features in the inputs does not lead to a significant increase in the AUC. For the INSIGHT and ADNI-MCI cohorts, it does lead to non-significant increase in the AUC, but the resulting cost for recruiting 100 individuals is higher (for INSIGHT, 527,437 € $\pm 36,332$, instead of 291,325€ $\pm 57,400$), as the cost of doing an MRI to each recruited individual has to be added to the initial cost. For ADNI-CN including MRI features in the input leads to a significant decrease in the AUC ($p < 0.01$). In all the cohorts, including MRI features leads to an increase in cost.

	INSIGHT cohort	ADNI-CN cohort	ADNI-MCI cohort
Proposed approach	67.5 (5.5)	69.1 (4.0)	82.4 (2.8)
MRI features only	61.9 (6.5)	59.0 (4.6)	80.1 (3.0)
MRI & cognitive features	68.8 (4.4)	67.1 (3.8)	82.8 (2.2)
With longitudinal variations	NA	71.7 (8.3)	87.7 (4.8)
After correction for age	68.5 (5.0)	67.7 (3.9)	80.9 (2.4)
ApoE4 only	63.7 (4.6)	62.1 (3.5)	75.1 (2.9)

TABLE 3.2 – Results in different experimental conditions. Data are : average percentage of Area Under the ROC Curve (standard deviation). NA = Not Applicable

3.4.2 Use of longitudinal measurements

Longitudinal measurements are available for individuals in the two ADNI cohorts. In order to evaluate the impact of using longitudinal measurements in amyloidosis prediction, the rate of change of the cognitive scores, computed using a 12-month visit, are included in the input features. The results, presented in line 4 of Table 3.2, show that the AUC is significantly better than the one obtained using only socio-demographic information, ApoE4 and cognitive scores at baseline, ADNI-MCI ($p < 0.0001$), and not significantly better for ADNI-CN ($p = 0.06$). Using longitudinal information overall leads to a better prediction.

However the cost of collecting such measurements has to be taken into account, since all individuals have to undergo cognitive assessments twice. Setting the cost of cognitive assessments for the second visit to 50€ for each individual, the total cost of recruiting 100 individuals using longitudinal information is of 243,448€ ($\pm 104,597$) for ADNI-CN and 133,452€ ($\pm 22,140$) for ADNI-MCI. This new cost is slightly lower than the one obtained using cross-sectional measurements in ADNI-CN ($234,591 \pm 23,106$) and higher for ADNI-MCI ($136,205 \pm 3678$). Therefore, although using longitudinal measurements leads to an increase in AUC, it does not lead to a decrease in recruitment cost.

3.4.3 Proposed method performance

3.4.3.1 Cost reduction

Table 3.3 presents the cost of recruiting 100 $A\beta+$ individuals in the different cohorts with the proposed method, as well as an estimation of the costs of recruiting

Dataset	Current method	Proposed method				
	Estimated current cost in €	% of AUC (std)	Individuals to be recruited	Number needed to scan	New cost in € (std)	Estimated savings in €
INSIGHT (27.7% A β +))	397,111 (N=361)	67.5 (5.5)	832	208	291,325 (57,400)	106,174
ADNI-CN (37.6% A β +))	292,553 (N=266)	69.1 (4.0)	599	175	234,591 (23,106)	58,063
ADNI-MCI (62.9% A β +))	174,880 (N=159)	83.8 (2.1)	264	112	138,294 (4857)	36586

TABLE 3.3 – Comparison of the proposed method results with the estimated initial costs for recruiting $K=100$ amyloid positive individuals. AUC = Area Under the ROC curve; std = standard deviation.

these individuals with the current method, consisting in scanning all potential individuals. This estimated current cost depends on the proportion of A β + in the data set. In order to find 100 A β + individuals in the INSIGHT cohort for example, $100/0.277 = 361$ individuals on average should be recruited and undergo a PET scan, which corresponds to a total cost of 397,111€. However, with the proposed method, about 832 individuals should be recruited and 208 PET scans would have to be done, leading to a cost of 291,325€ on average for recruiting 100 A β + individuals. The resulting savings would reach 106,174€ for this cohort.

The results presented in Table 3.3 show that the proposed method leads to a significant cost reduction when recruiting 100 individuals for all cohorts ($p < 0.001$), representing estimated savings of about 20%.

3.4.3.2 Age difference between groups

In the cohorts we used, the A β + individuals are older than the A β - individuals, especially in the ADNI cohorts (see Table B.1 in Supplementary Materials). One can therefore ask if the predictor is using this age difference, by simply predicting that older individuals are A β + and younger individuals are A β -, or by predicting the age of the individuals rather than their amyloid status. To confirm that it is not the case, we correct all the cognitive variables for age by using a linear regression and remove the age from the input features. After correction (results shown in line 5 of Table 3.2), the prediction performance is not impacted in INSIGHT and

Data set	AUC in %
Trained and tested on INSIGHT	61.9 (6.5)
Trained on ADNI-CN, tested on INSIGHT	62.0 (6.6)
Trained on ADNI-CN, tested on INSIGHT (all samples)	66.1 (3.6)
Trained and tested on [INSIGHT ADNI-CN]	61.3 (6.9)
Trained and tested on [INSIGHT ADNI-CN] (all samples)	67.5 (3.2)

TABLE 3.4 – Results using MRI variables, socio-demographic and genetic information on different data sets. Data are : average Area Under the ROC curve in % (std)

does not decrease significantly for ADNI-CN ($p>0.05$). In the ADNI-MCI cohort, correcting for age leads to a significant decrease in AUC ($p<0.01$) but results in a recruitment cost that is still significantly higher than doing a PET scan for all individuals ($p<0.01$). These results show that the prediction algorithm does not rely on the age difference between the groups and captures differences between amyloid positive and negative individuals that is not due to aging.

3.4.3.3 Training on a cohort and testing on a different one

The previous results are obtained by training and testing the method on distinct individuals from the same cohort. We want to confirm that these results would generalize well in a different setting, by verifying that they hold when the method is trained on a first cohort and tested on a different one.

ADNI and INSIGHT-preAD are very different studies. They have been designed for different purposes, as INSIGHT aims at studying very early phases of AD by studying changes appearing in healthy individuals, and ADNI aims at defining the progression of Alzheimer’s disease. The INSIGHT and ADNI-CN cohorts both include individuals who show no sign of dementia but with different inclusion criteria, and hippocampal measures have been extracted using different softwares. Hence, although these 2 cohorts can be compared, they are very different by design and purpose. In an ideal setup, cognitive features, socio-demographic information and ApoE4 should be used as input, however the cognitive assessments are different for ADNI and the INSIGHT-preAD study, hence they can’t be used as inputs when using these two cohorts.

We therefore train the prediction algorithm on ADNI-CN using socio-demographic information, ApoE4 and MRI features. We then test on INSIGHT the method trained on ADNI-CN in order to evaluate the generalization performance of our method. As the number of MRI features is large, LASSO feature selection was performed to select 12 MRI features. In order to have a fair comparison with training and testing on INSIGHT, the size of the selected training and test size are kept

the same as the training and test set coming from INSIGHT. We therefore randomly select $318 * 0.7 = 223$ from the ADNI-CN cohort to form the training set, and $318 * 0.3 = 95$ from INSIGHT to form the test set. This operation, followed by the classification, is performed 50 times in order to get a mean performance and a standard deviation.

The results, presented in Table 3.4, show that training on ADNI-CN and testing on INSIGHT gives similar performances to training and testing on the INSIGHT cohort.

3.4.3.4 Representativity of the selected population

For the selected individuals to be used as a clinical trial cohort, it is important to ensure that the selected population will be representative of the whole population of $A\beta+$ individuals that could have been selected. We therefore compare the individuals selected using the prediction method followed by a confirmatory PET scan with the $A\beta+$ individuals of the cohort.

We first pool together the test data set of the 50 cross-validation runs and look at the distribution of age, ADASCog (for ADNI cohorts), MMSE, education, age and gender. The histograms obtained for ADNI-CN are presented in figure 3.4. We can see that these histogram are very similar for age, gender, education, and cognitive features, but the proportion of ApoE4 carriers is higher in the group selected with the proposed method. Similar observations can be made for all cohorts.

In order to evaluate if there is a significant difference for each of these features, we compare the selected populations of the 50 runs with the populations of $A\beta+$ individuals of the corresponding test sets. A statistical test is performed for each of the 50 runs and a p-value is obtained for each of them. The used statistical test is a t-test for the features with a normal distribution (age and ADASCog), a binomial proportion test for binary features (presence of ApoE4 alleles and gender) and a Mann–Whitney U test for the remaining features (MMSE and education). A p-value is obtained for each run, for each feature. Figure 3.5 presents the proportion of these p-value that are below 0.05, for each feature.

The main bias that can be seen across cohorts is a higher proportion of ApoE4 carriers, which is statistically significant in 16% of cases for INSIGHT, 48% for ADNI-CN and 98% for ADNI-MCI. Although this bias is important, especially for the ADNI cohorts, it seems acceptable as many current recruiting procedures also have this bias or only recruit ApoE4 carriers, such as in the Alzheimer’s Prevention Initiative Generation study (LOPEZ et al., 2017).

The proposed method leads to an unbiased cohort in terms of age, gender, and education, as well as cognitive scores in more than 94% of cases for the asymptomatic cohorts, and 82% for ADNI-MCI.

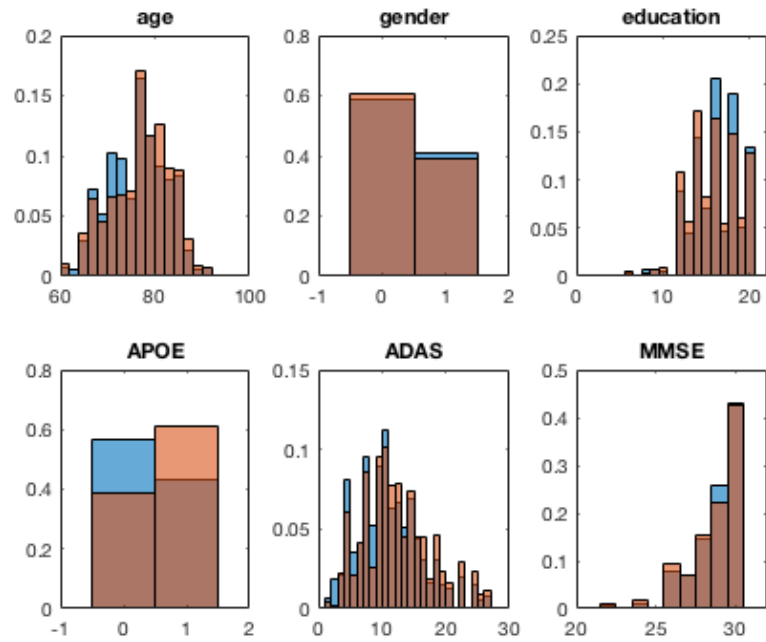


FIGURE 3.4 – Histogram of the different features for the selected group (orange) and for the whole $A\beta+$ group (blue), for the ADNI-CN cohort

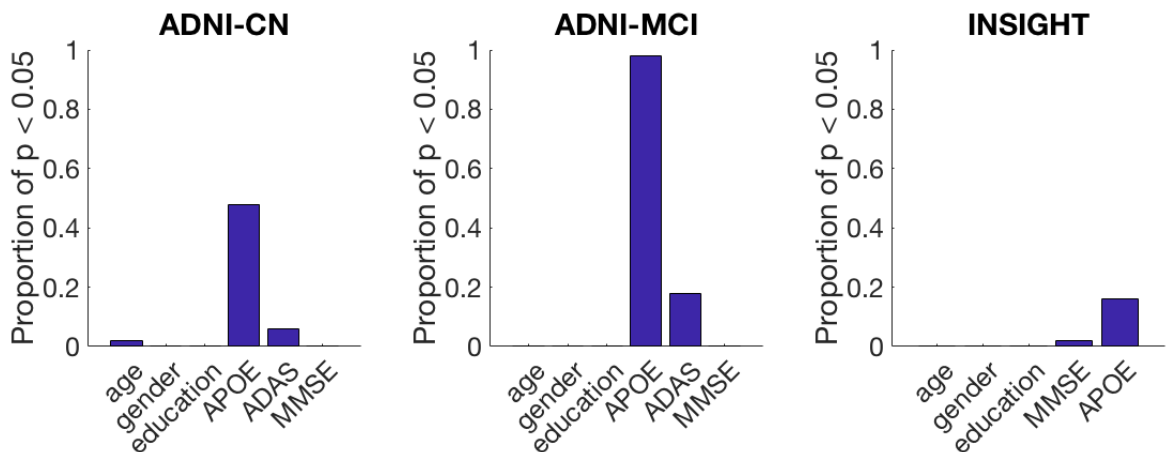


FIGURE 3.5 – Proportion of runs with a significant difference between the groups for each feature, in each of the 3 cohorts

3.4.4 Building larger cohorts

3.4.4.1 Pooling data sets

Different cohorts can be pooled in order to create a bigger data set, containing a large number of individuals. However, this operation requires that the heterogeneity of the pooled cohort does not alter the performances of the method that is applied. In order to verify this hypothesis, we pool the ADNI-CN cohort with the INSIGHT cohort. We train and test the method on individuals coming from both of this cohort, using the same training and test size as in INSIGHT, in order to compare the performances with the one obtained by training and testing solely on INSIGHT. As in the generalization experiment, we use MRI features instead of cognitive features which are different in the 2 cohorts. The results, presented in Table 3.4 show that the performances are not significantly different when the algorithm is trained and tested on the pooled cohort, which shows that the heterogeneity of pooled data sets does not alter the classification performances.

3.4.4.2 Effect of sample size

When learning on ADNI-CN and testing on INSIGHT to test generalization, we used the same training and learning size as in INSIGHT to have a fair comparison, hence using only 52% of the available data at each run. For the same reason, we used only 42% of the created cohort when we pooled the INSIGHT and the ADNI-CN cohort. We now want to measure the impact of increasing the cohort size by using the full cohort in each case, always keeping the same ratio for the size of the training and test data sets (70%-30%). The results, presented in Table 3.4 show that increasing the cohort size significantly increases the performances ($p < 0.0005$). This result comforts the need to create large data sets, or pool existing ones, to create more accurate prediction tools.

3.5 Discussion

3.5.1 Results of the experiments

3.5.1.1 Algorithm and feature choice

The algorithm benchmark shows there is not one outstanding algorithm that would outperform all the others on all data sets. These findings support the "No free lunch" theorem (WOLPERT et MACREADY, 1997; WOLPERT, 2002), stating that different algorithms perform best on different problems. As a choice had to be made, we used the Random Forest which performed well on the 3 cohorts. It is

not however a general recommendation. When working on a new classification problem, even similar to this one, one should always compare different algorithms to choose the most suited one.

Because the number of features is large compared to the number of available subjects, using all the available features may result in a low performance (HUGHES, 1968). The low performance we obtained on the INSIGHT cohort using all the available cognitive features is an illustration of this phenomenon, known as the curse of dimensionality. A typical way of solving this issue is using automatic methods for dimension reduction. We showed that, in our case, selecting features using expert knowledge gives better results. It corroborates the fact that when a large number of features and a small data set are available, feature engineering using domain knowledge is necessary (DOMINGOS, 2012).

Hypothetical models of AD suggest neurodegeneration and changes in structural MRI appear earlier than cognitive decline (JACK, KNOPMAN et al., 2010). This hypothesis is supported by findings from Bateman et al. (BATEMAN et al., 2012), showing that, in autosomal dominant AD, brain atrophy occurs 15 years before AD diagnosis, 5 years before episodic memory decline and 10 years before changes in other cognitive domains. Studies by Ameiva et al. show changes in several domain of cognition can be observed 9 years before diagnosis (AMIEVA, JACQMIN-GADDA et al., 2005), and up to 16 years before diagnosis for individuals with higher education (AMIEVA, MOKRI et al., 2014). Overall, brain atrophy may appear before or at about the same time as cognitive decline, and one could expect using MRI would improve the prediction of amyloidosis, especially for cognitively normal individuals. Our analysis however suggests that it is not the case. This finding that clinical signs can allow for efficient pre-screening goes against the current purely biological definition of AD by NIA-AA (JACK, BENNETT et al., 2018). We can suppose memory decline has already started for individuals with a SMC, so that cognitive features are already slightly altered. It leads us to think that subtle cognitive changes appear in late preclinical AD, as hypothesized by Sperling et al. in their 3 stage model of pre-clinical AD (SPERLING et al., 2011). The results can however depend on the choice of MRI features. In future studies, different neuroimaging features could be used to test this hypothesis that cognitive changes are anterior to substantial structural changes, in line with previous studies on optimal neuroimaging feature selection in pre-clinical AD (JACK, WISTE et al., 2015). Alternatively, a more advanced feature selection algorithm might be able to identify the most informative MRI features and therefore improve their performance, as proposed in other methods (KATE et al., 2018).

In the ADNI-CN cohort, adding the MRI features even leads to a decrease in AUC, whereas it leads to a slight increase for INSIGHT. A possible explanation for

this difference between cohorts is that in ADNI, the number of cognitive features (4) is low compared to the number of MRI features (73), whereas the difference is smaller for INSIGHT (26 cognitive features for the same number of MRI features). In ADNI the cognitive scores can therefore be under-represented compared to the MRI features. This effect should be handled by the Random Forest, that can give different weights to different features. It however requires the number of individuals to be large enough compared to the number of features, which is not the case here.

Overall, we showed that with our method the best results are obtained without performing an MRI and without longitudinal features, but using only data that can be easily acquired. MRI should not be performed in the pre-screening phase, however performing an MRI at the end of the recruitment process will always be needed to exclude vascular lesions or tumors and as a reference for adverse event monitoring.

3.5.1.2 Method performance

We showed that using the proposed method as a pre-screening phase for individual recruitment in clinical trials leads to reducing the recruitment cost by about 20%. These findings are however based on cost hypothesis that can seem arbitrary. In particular, the cost of recruiting a new subject is the same whatever the number of subjects that have been recruited. In practice, because a large number of studies intend to recruit large numbers of subjects, the more subjects are recruited, the more difficult it is to recruit a new one. Having a non-constant cost could therefore represent an improvement of the proposed method and be closer to the difficulties encountered in practice.

We can expect the method to generalize well and give similar results when applied on any cohort of cognitively normal individuals because we showed we obtain similar performances when training and testing on the same cohort or on two different ones. The cohorts we used for testing are slightly unbalanced, with $A\beta+$ individuals older than $A\beta-$ individuals, but correcting for age gives similar cost reductions, so the same results should be obtained on cohorts that do not have the same unbalance. Comparing the selected $A\beta+$ individuals with all the $A\beta+$ individuals of the cohort shows that the subset selected with the proposed method is unbiased. The proposed method therefore leads to the recruitment of a representative cohort with a reduced cost.

The proposed approach is time efficient, as in the worst case the training phase may take few minutes, while testing a new subject could be done in less than a

second. Therefore, computational time is not a limiting factor for using such methods in practice. Furthermore, since only clinical data may be used for good performance, the method could be easily deployed in the current clinical practice.

3.5.1.3 Data set size

Table 3.4 shows that pooling data sets does not alter the performance of the prediction, although it brings heterogeneity; and that increasing the cohort size improves the prediction. This last finding is supported by the current machine learning literature, stating that gathering more data often yields an increase in performance greater than the increase one could obtain by improving the prediction algorithm (DOMINGOS, 2012). It shows the importance of gathering more data in the medical field and more specifically related to dementia. While the largest cohorts widely available usually include less than 1500 subjects, creating larger cohorts could result in a significant increase of performance for predicting amyloidosis or for other predictive task, such as automatic diagnosis based on neuroimages (FRANKE et al., 2010; ARBABSHIRANI et al., 2017). As long as larger cohorts are not available, we recommend pooling different cohorts in order to get a better prediction performance. For example, the preclinical cohorts presented by Epelbaum et al (EPELBAUM, GENTHON et al., 2017) could be pooled to create a bigger cohort to train and validate our method.

3.5.2 Comparison with existing methods

3.5.2.1 Univariate approaches

A standard approach for prediction is using univariate methods. As a comparison with our method, a Random Forest is trained and tested on each input variable separately. The best univariate results are obtained using ApoE4 (Table 3.2, line 4). The AUC obtained using ApoE4 is significantly lower ($p < 0.0001$) than the AUC of the proposed multivariate method, for all cohorts, with an AUC of 63.7 ± 4.6 instead of 67.5 ± 5.5 for INSIGHT for example. The proposed method therefore outperforms its univariate equivalent.

3.5.2.2 Other multivariate approaches

We wanted to compare the performance of our method with that of other similar studies. Different cohorts and different performance measures have been used in these studies, the comparison is therefore not straightforward and the results should be interpreted with caution.

In the study of Mielke et al (MIELKE et al., 2012) the studied cohort is composed of CN individuals from the Mayo Clinic Study of Aging. This cohort is comparable with the ADNI-CN cohort used in this work, as individuals from both cohorts are CN, and the ratio of $A\beta+$ individuals is close (34.9% in the Mayo Clinic Study of Aging cohort, 37.6% in ADNI-CN). A logistic regression is used with an a priori set and non-optimized threshold, and the performance measures were obtained by training and testing the algorithm on the same individuals. The resulting AUC, of 0.71, is significantly better than the AUC we obtain on ADNI-CN (69.1, $p < 0.05$), which is expected as training and testing an algorithm on the same individuals generally gives better results than testing it on a different set of individuals.

The cohort used by Insel et al (INSEL et al., 2016) contains CN individuals, with a proportion of positive individuals of 40.8%, so the closest cohort is again ADNI-CN. The AUC is not provided in the study, so it cannot be used for comparison. The Positive Prediction Rate (PPR) and Negative Prediction Rate (NPR) are however given and, as shown in Supplementary Materials, they can be used to compute S and R. The normalized cost can therefore be computed, and is significantly lower ($p < 0.0001$) with our method.

The AUC we obtain on the MCI cohort is comparable to the ones obtained in other studies or slightly higher (TOSUN, JOSHI et WEINER, 2013; APOSTOLOVA et al., 2015; KATE et al., 2018). Ten Kate et al (KATE et al., 2018) obtain a slightly better AUC for the prediction in CN subjects. This difference might be explained by the use of a different feature selection method.

3.6 Conclusion

We proposed a method for creating cohorts of $A\beta+$ individuals with a reduced recruitment cost. In a pre-screening phase, we use a classifier to identify a sub-population of individuals who are more likely to be amyloid positive, based on clinical data. We then do a confirmatory PET scan to the individuals of this sub-population only. The whole algorithm has been optimized so as to minimize the cost of the cohort recruitment. As such automatic methods are today limited by the number of subjects, future studies could be performed on a Phase 3 clinical trial cohort, as such cohorts often include more than 1000 participants. New screening technologies, such as blood-based biomarkers (SHAW et al., 2009; NAKAMURA et al., 2018), could transform the recruitment process for clinical trials, which could also be facilitated by web-based cognition evaluation systems, such as the Brain Health Registry (trial identifier : NCT02402426).

Chapitre 4

Use of psychotropic drugs throughout the course of Alzheimer's disease : a large-scale study of French medical records

This chapter is in preparation for submission as a journal article.

4.1 Introduction

Recent therapeutic trial interruptions in the field of Alzheimer's disease (AD) have been a tremendous disappointment for patients, their families and the scientific and medical communities alike. There is a critical need to upgrade our understanding of modifiable risk factors leading to this devastating disease for primary prevention purposes (NORTON et al., 2014). To achieve this goal, many studies have focused for instance on vascular risk factors (WHITMER et al., 2005), psychiatric illnesses (BARNES et YAFFE, 2011) and psychotropic drug intake (BILLIOTI DE GAGE et al., 2014; BIÉTRY et al., 2017; S. L. GRAY et al., 2016). However, multiple risk factors have seldom been assessed simultaneously although it seems that multimorbidity, that is the co-occurrence of at least 3 diseases (MARENGONI et al., 2009) is associated to AD neuroimaging makers even at the preclinical stage (MENDES et al., 2018). These risk factors can help in future trials as enrichment inclusion criteria. They may also yield insights into the aetiopathogeny of AD. Finally, their identification can help to provide successful prevention strategies. Previous studies, such as (MENDES et al., 2018; ANSART, EPELBAUM, GAGLIARDI, COLLIOT, DORMONT, DUBOIS et al., 2019) acknowledge generalizability limitations due to the small sample size. However, the availability of larger and larger databases of health records now facilitates analyses on large general population samples which allows to better identify chronic diseases risk factors (PERERA et al., 2014;

W.-Y. LIN et al., 2019). In this study, we analyze the medical records from more than 60,000 individuals using a standardized digital database called Cegedim.

4.2 Materials and methods

4.2.1 Cohort description

4.2.1.1 Description

Cegedim is a company developing and commercializing health management software (standardized electronic record files), hence gathering data on patient follow-up in the health care system. Its products are used by 25 000 health practitioners in France, among which 3000 have been recruited to constitute GERS-DATA, also known as THIN (The Health Improvement Network). This observatory includes 2000 general practitioners, which have been used for this study, and 1000 specialists. These practitioners have been selected so as to be representative of the global practitioner cohort in terms of gender, age and geographic position. All the prescriptions made by these practitioners are paired with a corresponding prescription diagnosis.

The collected data is fully anonymized so as to be General Data Protection Regulation (GDPR) compliant.

4.2.1.2 Group definition

Three groups have been defined :

- The AD group includes patients diagnosed with Alzheimer’s disease dementia (international classification of diseases 10th edition : ICD10 codes F00 or G30), that have been followed for at least 2 years before this first diagnosis. Patients diagnosed with AD before being 50 years old have been excluded from the study.
- The MCI group includes patients diagnosed with a memory impairment (ICD10 codes F06.7 or R41) that is not explained by any neuro-degenerative conditions. This cohort has been matched for age and sex with the AD cohort. Complete list of exclusion diagnosis : dementia (F00-F03), mental retardation (F70–F79), disorders of psychological development (F80–F89), inflammatory diseases of the central nervous system (G00–G09), systemic atrophies primarily affecting the central nervous system (G10–G13), extra-pyramidal and movement disorders (G20–G26), other degenerative diseases of the nervous system (G30–G32), demyelinating diseases of the central

nervous system (G35–G37), epilepsy (G40–G42), cerebrovascular disorders (G45–G46).

- The CN group includes patients with no ICD10 diagnosis of category F (Mental and behavioral disorders) or G (Diseases of the nervous system), matching the AD cohort for age and sex. As many patients in France fulfill these criteria, only randomly selected age and gender matched patients followed for at least 7 years have been included so as to have a similar number of patients as in the AD group.

4.2.1.3 Patient overview

A description of the 3 groups is shown in Table 4.1.

	AD	MCI	CN
number of patients	22 272	12 334	25 956
Age group (%)			
21-50	6.2	0.5 ***	0.5 ***
51-75	54.6	48.7 ***	51.2 ***
>75	39.2	50.8 ***	48.3 ***
Gender (%)			
Male	35.9	35.7	35.8 ***
Female	64.1	64.3	64.2 ***
Number of visits / patient	67.53 (56.89)	47.89 (54.51) ***	49.20 (47.55) ***
Number of days between 2 visits	57.42 (133.60)	57.90 (137.28) ***	94.44 (280.09) ***
Follow-up interval in years	10.46(4.84)	7.43(6.04) **	12.46(4.03) ***

TABLE 4.1 – Cohort description. Data are mean (standard deviation). * = significant at the 0.05 level; ** = significant at the 0.01 level; *** = significant at the 0.001 level (two-sided t-test with Bonferroni correction for multiple comparison).

4.2.2 Studied treatments

When a prescription is made by the practitioner using the Cegedim software, the treatments listed on the prescription are automatically added to the database. Studying treatment instead of diagnosis manually added by the clinician is therefore reduces variability due to the clinician usage of the software.

It is to be noted that this list of prescribed treatments is available only for prescriptions made by the general practitioner following each patient, prescriptions made by other practitioners are therefore not available.

The studied treatment categories are defined according to the ATC codes as follows :

- glucose lowering treatments (A10A, A10B)
- tension reducing treatments (C02, C03, C07, C08, C09)
- anti-inflammatory and anti-rheumatic treatments (M01)
- anti-psychotic treatments (N05A)
- benzodiazepine (N05BA, N05CD, N05CF)
- antidepressant (N06A)
- dementia drugs (N06D)
- herpes treatments (J05AB01, J05AB09, J05AB11)

We chose to interest ourselves in broad categories of treatment instead of individual molecules to derive more global messages from our findings. The choice of categories was based on a review of the literature.

For each treatment category and for each subject at each semester, we create a feature of values 1 if the subject has been prescribed a treatment of the given category at least once during the semester.

For patients from the AD group, time 0 corresponds to the semester of first AD diagnosis. For patients from the MCI and CN groups, time 0 corresponds to the time at which the subject is 80 years old, which is the median age of AD diagnosis in the AD cohort. This choice allows to compare the treatment evolution in different groups at similar ages.

4.2.3 Descriptive and predictive analysis of treatment history

4.2.3.1 Statistical analysis

We perform two group comparisons : AD vs. MCI and AD vs. CN. For each comparison, we consider the log-odds of being treated with a category of drugs in the two groups for each semester of the total follow-up period of 25 years. We model the change of these log-odds with time using a generalized mixed effect model with logit as link function and the outcome being the presence of a prescription for each subject at each semester (see Supplementary Materials for details). In the AD group, the model assumes a different linear change before and after diagnosis; both linear functions have a fixed intercept and slope, and a random intercept is added for each subject. In the other groups, the model assumes a single linear function with a fixed intercept and slope and a random intercept.

We then test whether slopes and intercepts are statistically different in the pre-diagnosis period between both groups. We also test the change in slope and intercept between the pre-diagnosis and post-diagnosis period within the AD population. We use Wald tests corrected for multiple comparisons using the Bonferroni

method with a significance threshold of 5%.

This analysis was performed in R, using the `glmer` function of the `lme4` package.

4.2.4 Predictive model

We use a machine learning approach to predict if an individual will have a diagnosis of Alzheimer's disease in the next 5 or 10 years based on the treatments of the individual in a given semester. A positive case is therefore an individual from the AD cohort who is not diagnosed with AD at the considered semester and has been diagnosed with AD within the following 5 or 10 years.

For all individuals in the AD cohort, a semester before AD diagnosis is randomly selected, which avoids repeated data. If the selected semester is followed by an AD diagnosis in the next 5 or 10 years, the individual is attributed to the set of positive cases. Otherwise, it falls within the set of negative cases (see Scenario 1 below).

We evaluate different scenarios depending on the definition of a negative case :

- Scenario 1 : A negative is an individual from the AD cohort, who is not diagnosed with AD at the considered semester, have follow-up data until 5 years, and had not been diagnosed with AD within this time period.
- Scenario 2 : A negative is an individual who satisfies the previous definition at a current semester or an individual from the MCI cohort at the first available semester who had been followed for 5 years (therefore without AD diagnosis). This definition adds 8,412 negative cases from the MCI cohort to the scenario 1.
- Scenario 3 : A negative is as in the scenario 2 but with MCI replaced by CN. It adds 29,513 negative cases from the CN cohort to the scenario 1.
- Scenario 4 : A negative is as defined in the scenario 1 but with a 10 years follow-up period instead of 5.

Figure 4.1 shows an overview of the individuals included in the positive and negative case sets for each scenario.

Scenario 1 is difficult since an individual progressing to AD at 5 years and 6 months will be considered negative whereas it is very close to be positive. Scenario 4 aims to alleviate the threshold effect by increasing the time-period, at the cost of reducing the interest of the method for the detection of patients at-risk of rapid progression to AD. When selecting a random visit before diagnosis, an average of 69.2% (± 0.23 std) of patients have a AD diagnosis in the next 5 years, and 92.1% (± 0.15 std) do in the next 10 years.

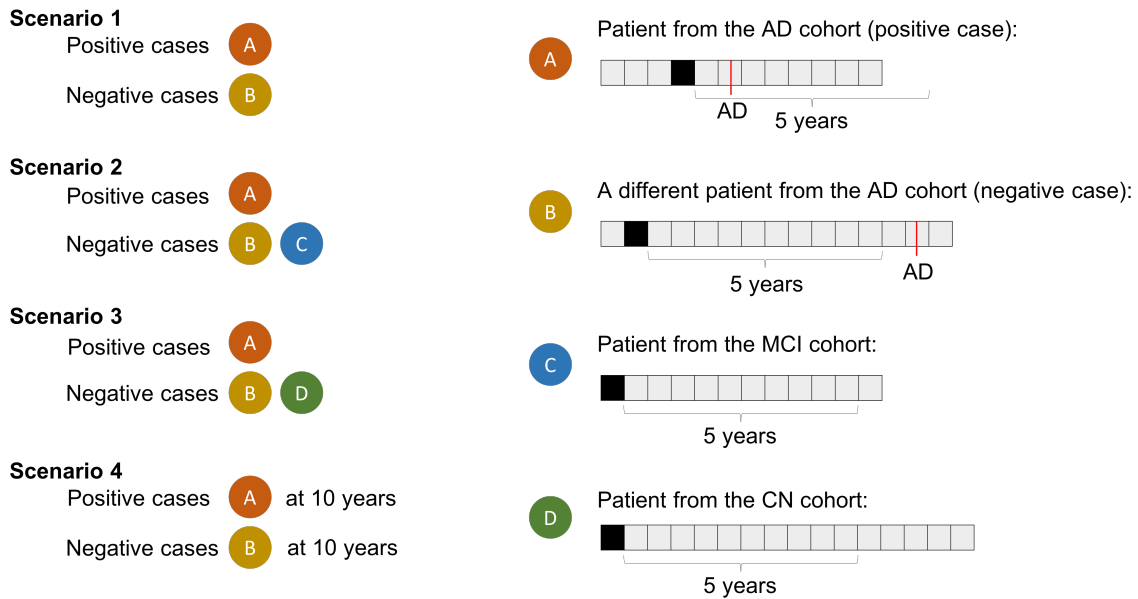


FIGURE 4.1 – Positive and negative cases for the different scenarios of the predictive model

Nevertheless, both scenarios do not mimic the real clinical practice, since the negative cases are chosen within the AD cohort, therefore knowing beforehand that the individuals will eventually develop AD. Scenario 2 is the one who would generalize best in the real life, since negative cases are any MCI subjects at present. It increases the risk of false positive compared to scenario 1.

Age, gender, and the treatment categories (for each category, the prescription of at least one treatment in the last 6 months) were used as input in a logistic regression classifier. The performance is evaluated performing a 5-fold cross-validation. 10 repeats of the procedure with random split between train and test sets yield and average and standard deviation of performance measures over a total of 50 experiments. Performance measures include the area under the receiver operating characteristic (ROC) curve (AUC), and for a given threshold : sensitivity (sen), specificity (spe), accuracy (acc) and balanced accuracy (bacc). We used Youden's method to choose the point on the ROC curve, hence maximizing Youden's J statistic (J), which is equivalent to maximizing the balanced accuracy. In an additional experiment, we choose the point on the ROC curve so as to maximize the sensitivity while keeping a specificity of at least 80%. These additional results are shown in Annexe C.

4.3 Results

We identified 60,730 patients in the Cegedim database which characteristics are described in Table 4.1. Using a logistic mixed effect model (Figure 4.3 and 4.2

and Table 4.2 and 4.3) we show that drug prescriptions for patients diagnosed with AD are higher before diagnosis for antidepressant, antipsychotic and anti-dementia drugs compared to MCI and for all studied drug categories when compared to controls (with the highest odd ratios obtained for the same psychotropic drugs as in the AD vs. MCI + benzodiazepine in the AD vs. CN comparison). Secondly, when we consider the slopes of prescriptions before AD diagnosis and compare them to that of the MCI and CN groups, we evidence differences suggesting dynamic changes across the AD continuum. Looking at the AD vs. MCI or CN models, the most striking differences observed both before and after AD diagnosis were observed for psychotropic drugs and especially for antidementia and antipsychotic drugs. At the time of diagnosis there is a dramatic increase in antidementia, antipsychotic and antidepressant drugs in the AD group compared to just before diagnosis while the prescription of other drug categories (except for anti-herpetic drugs) are decreased, most notably for anti-inflammatory and anti-rheumatic drugs. Then, interestingly, we evidence a gradual decline in the usage of all type of drug (including antidementia drugs) to the exception of anti-herpetic drugs in the years following AD diagnosis compared to the prescription practices in our two control groups.

4.4 Discussion

In this large sample representative of the general population seen in general practitioner offices in the last 25 years in France we evidenced different prescription practices in patients with AD diagnosis as compared to patients with stable MCI and normal cognition. This case-control study benefits from a large-scale clinical database, called Cegedim, that has been anonymized and deidentified for clinical research purposes. Among our findings we can distinguish two different domains : firstly, in the period preceding the diagnosis of AD we can identify probable risk factors and secondly, in the period encompassing the time of diagnosis and afterwards we can analyze the drug related management of AD.

4.4.1 Risk factors

We interested ourselves to different classes of reported risk factors of AD. For instance, infection by herpes simplex virus type 1 (HSV-1) (S. A. HARRIS et E. A. HARRIS, 2018). HSV-1 is indeed a neurotropic virus that is highly prevalent in the aged population. Both genomic and proteomic studies revealed an HSV-1 enrichment in AD brains. Epidemiological data have repeatedly confirmed the link between HSV-1 & AD. Genetic risk factors for AD (e.g. APOE4) also play a role

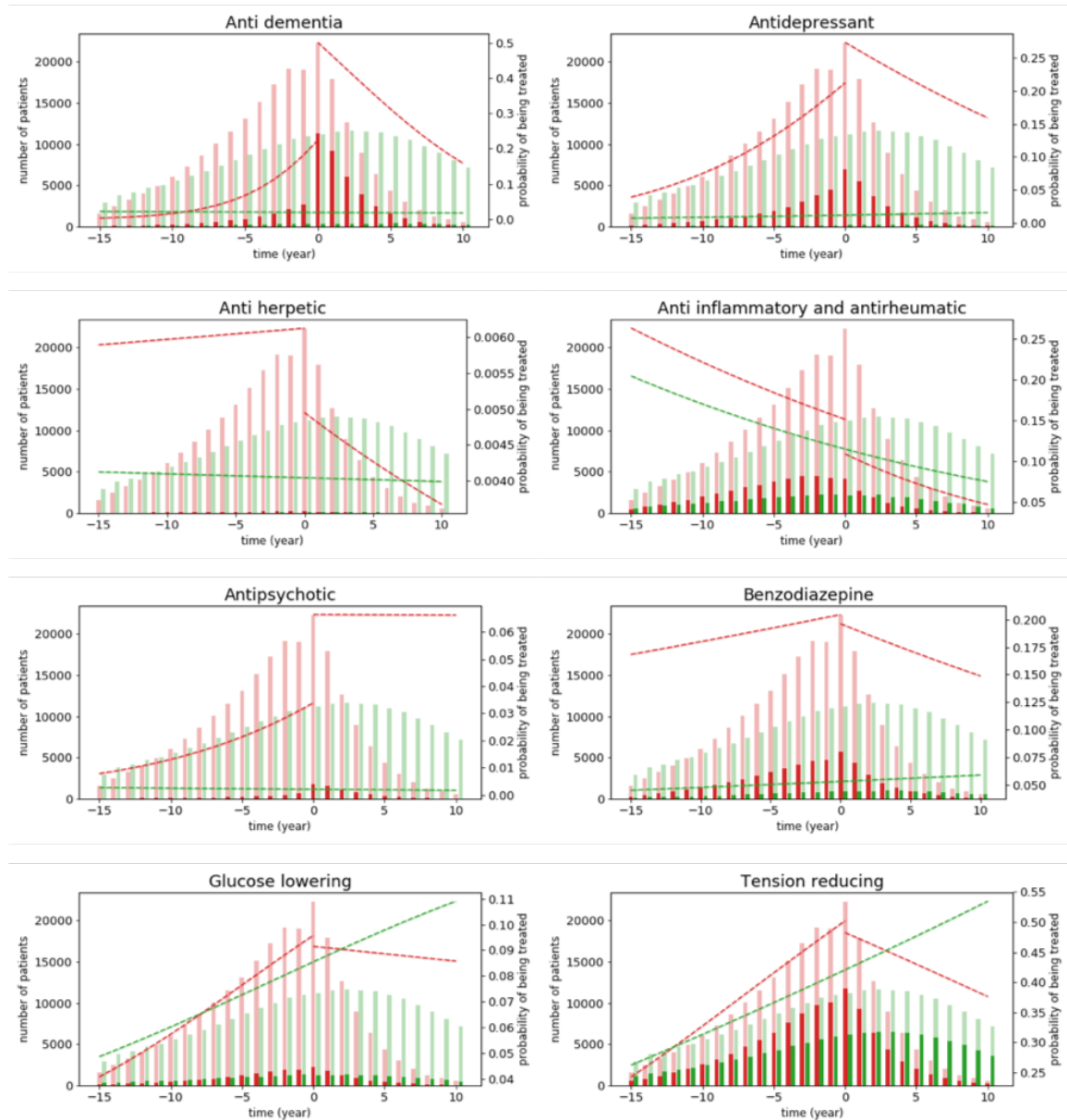


FIGURE 4.2 – Proportion of treated subjects such as observed in the cohort and such as fitted by the model, for the AD group (in red) and the CN group (in green). The bar plots show to the number of subjects observed at the corresponding time points and how many of them where treated (solid colors) or not (transparent colors) (y-axis on the left). The dashed lines correspond to the probability of being treated at each time point for the different groups, such as fitted by the model (y-axis on the right).

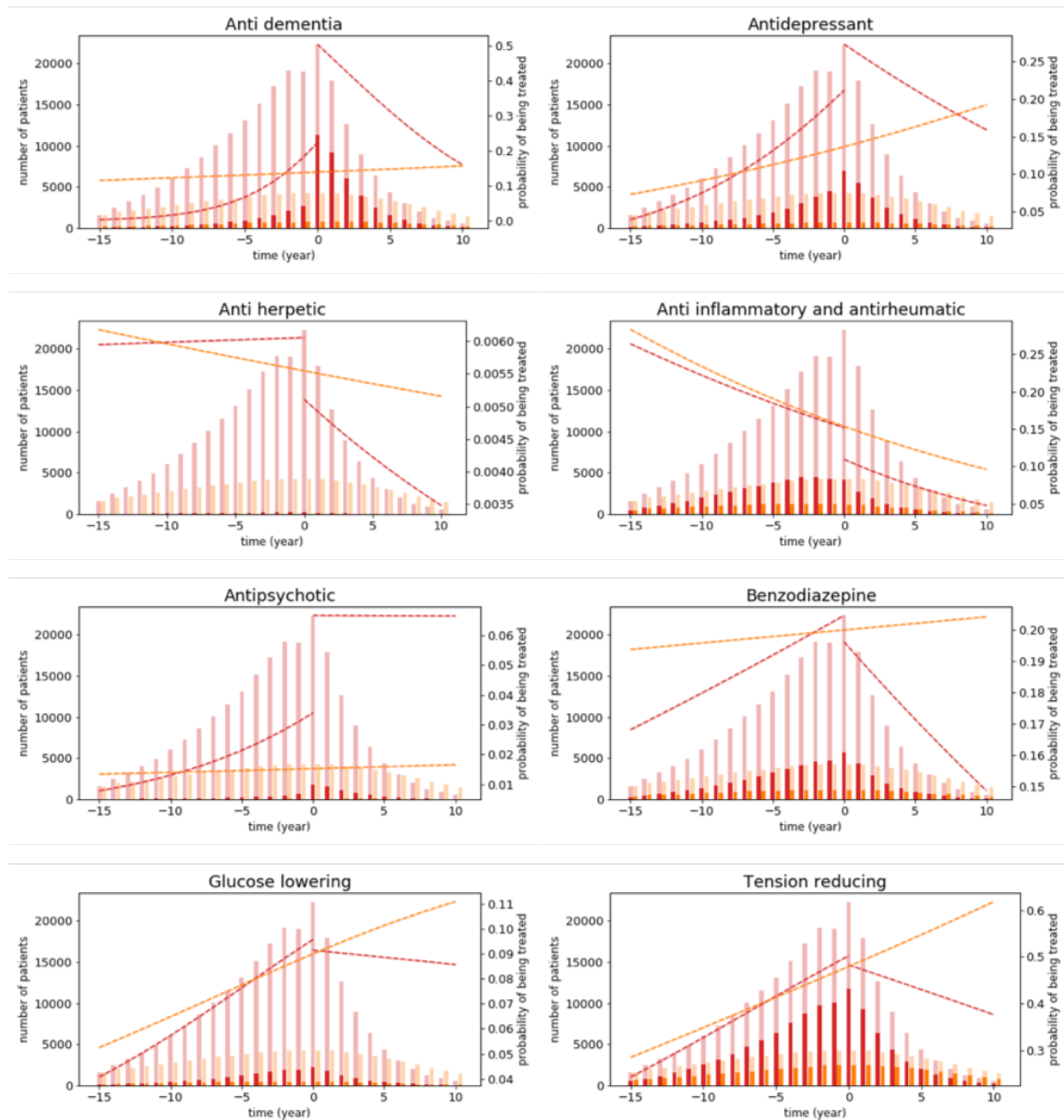


FIGURE 4.3 – Proportion of treated subjects such as observed in the cohort and such as fitted by the model, for the AD group (in red) and the MCI group (in orange). The bar plots show to the number of subjects observed at the corresponding time points and how many of them where treated (solid colors) or not (transparent colors) (y-axis on the left). The dashed lines correspond to the probability of being treated at each time point for the different groups, such as fitted by the model (y-axis on the right).

	Odd ratio of AD vs MCI before diagnosis	Ratio of slope for AD subjects vs MCI subjects	Odd ratio of after diagnosis vs before diagnosis for AD subjects	Ratio of slope after diagnosis vs before diagnostic for AD subjects
Anti herpetic	1 (0.11)	1.01 (0.011)	0.797 (0.056)	0.95 (0.022)
Anti inflammatory and anti-rheumatic	0.937 (0.024)	1.01 (0.0025)**	0.605 (0.0079)***	0.952 (0.0057)***
Antidepressant	2.76 (0.34)***	1.12 (0.0046)***	1.88 (0.071)***	0.722 (0.0031)***
Antipsychotic	2.39 (0.52)***	1.13 (0.011)***	3.21 (0.38)***	0.869 (0.0077)***
Benzodiazepine	1.11 (0.052)	1.02 (0.0032)***	0.906 (0.017)***	0.913 (0.005)***
Antidementia drugs	2.84 (0.29)***	1.45 (0.0088)***	7.54 (1.1)***	0.519 (0.0017)***
Glucose lowering	1.2 (0.16)	1.03 (0.0071)***	0.802 (0.027)***	0.8 (0.0071)***
Tension reducing	1.04 (0.051)	1.05 (0.0033)***	0.812 (0.013)***	0.75 (0.0031)***

TABLE 4.2 – Odd ratios of prescription practices before and after diagnosis of AD as compared to MCI control group. Data are odd ratios (Standard deviations). * = significant at the 0.05 level; ** = significant at the 0.01 level; *** = significant at the 0.001 level (Bonferroni correction for multiple comparison was applied). AD : Alzheimer's disease, MCI : mild cognitive impairment.

	Odd ratio of AD vs CN before diagnosis	Change of slope for AD subjects vs CN subjects	Odd ratio of after diagnosis vs before diagnosis for AD subjects	Change of slope after diagnosis for AD subjects
Anti herpetic	1.38 (0.12)***	1.01 (0.0069)	0.753 (0.051)*	0.958 (0.023)
Anti inflammatory and anti-rheumatic	1.55 (0.055)***	1 (0.002)	0.604 (0.0079)***	0.952 (0.0058)***
Antidepressant	107 (5.4e+02)***	1.16 (0.0061)***	1.89 (0.073)***	0.72 (0.0031)***
Antipsychotic	17.2 (35)***	1.19 (0.015)***	3.2 (0.37)***	0.87 (0.0077)***
Benzodiazepine	23.4 (23)***	1.01 (0.003)	0.904 (0.017)***	0.912 (0.005)***
Antidementia drugs	94.8 (3.7e+02)***	1.52 (0.011)***	7.97 (1.3)***	0.51 (0.0016)***
Glucose lowering	1.79 (0.28)***	1.05 (0.0055)***	0.802 (0.027)***	0.8 (0.0071)***
Tension reducing	2.06 (0.18)***	1.06 (0.0027)***	0.811 (0.013)***	0.749 (0.0032)***

TABLE 4.3 – Odd ratios of prescription practices before and after diagnosis of AD as compared to CN control group. Data are odd ratios (Standard deviations). * = significant at the 0.05 level; ** = significant at the 0.01 level; *** = significant at the 0.001 level (Bonferroni correction for multiple comparison was applied). AD : Alzheimer’s disease, CN : cognitively normal.

	AUC	J	acc	bacc	sen	spe
AD and MCI groups (5 years)	70.8 (0.61)	32.3 (1.2)	66.2 (0.6)	66.2 (0.6)	61.7 (3.1)	70.7 (3)
AD and CN groups (5 years)	70.5 (0.6)	30.5 (1.1)	67.8 (1.4)	65.3 (0.54)	58.9 (3.7)	71.6 (3.4)
AD group (5 years)	69.2 (0.85)	30.4 (1.5)	62.8 (1.6)	65.2 (0.74)	59 (3.8)	71.4 (4)
AD group (10 years)	75.6 (1.1)	41.7 (2.3)	63.7 (3.6)	70.9 (1.2)	62.4 (4.2)	79.3 (4.4)

TABLE 4.4 – Performance of the prediction of the presence of an AD diagnosis 5 or 10 years after a random visit for different groups of subjects. AUC = area under the receiver operating characteristic curve; J = Youden’s J statistic; acc = accuracy; bacc = balanced accuracy; sen = sensitivity; spe = specificity.

in the HSV-1 life cycle/infectivity. In vitro and in vivo, HSV-1 favors A β production as well as increased phosphorylation of Tau in neurons (CHIARA et al., 2019; MARTIN et al., 2014; WOZNIAK et al., 2007).

An estimated 3.7 billion people are today infected with HSV-1 (ORGANIZATION, 2017) and there is a 90% prevalence of the virus in populations after the age of 50 years. In a recent population study on 33.000 Taiwanese individuals (TZENG et al., 2018), the risk to develop AD was 2.5 fold greater in infected people with recurrent viral reactivations. This risk returned to baseline in people treated with antiviral medications. In our study, we evidence an increased prescription of anti herpetic drug prescription before time of AD diagnosis compared to the CN group that would support these claims.

Midlife diabetes (CHENG et al., 2012), and more generally, vascular risk factors (WHITMER et al., 2005), have been identified as dementia risk factors. Again, in our study, AD patients were more frequently treated with tension and glucose reducing drugs prior to diagnosis as compared to CN.

Anti-inflammatory and antirheumatic drugs were more frequently prescribed in AD patients before diagnosis as compared to the prescription frequency in the CN group. The relation between systemic inflammation and AD has been explored thoroughly in the last two decades (HOLMES, 2013) and recent findings support a role for peripheral inflammation as early as the prodromal stage of AD and dementia with Lewy Bodies (KING et al., 2018). Our finding suggests that this inflammation might be earlier still and indeed, another recent study has shown that neuroinflammation predates amyloid deposition in the brain of patients with prodromal AD (HAMELIN et al., 2016). At the time of diagnosis, the prescription frequency of this type of drugs falls below that of stable MCI and NC groups and continues to decrease afterwards. This is probably due to the rate of adverse events with non-steroidal anti-inflammatory (NSAID) drugs (HARIRFOROOSH, ASGHAR et JAMALI, 2013) especially in patients with cognitive decline who may experience treatment observance difficulties. Finally, the fact that the efficacy of aspirin, steroid and NSAIDs (traditional NSAIDs and selective cyclooxygenase-2 inhibitors) is not proven and thus not recommended for the treatment of AD (JATURAPATPORN et al., 2012) probably accounts for the findings after AD diagnosis in our study.

Finally, the most dramatic differences were evidenced for psychotropic drugs. There was a gradual increase in the over prescription of antidepressant, antipsychotic, and antidementia drugs in the 15 years preceding diagnosis. Interestingly, the probability of being treated by one of these drugs was already superior to

that of CN 15 years before diagnosis while it was inferior to that of MCI until 10 to 5 years before AD diagnosis and superior afterwards. As in any case-control study we can only hypothesize about such findings. Some authors have proposed that differences evidenced 15 years before AD diagnosis are indeed directional in the sense that it is hardly plausible that AD is already clinically relevant at this point to justify a psychotropic treatment (RICHARDSON et al., 2018). However, our prescription probability curves are really reminiscent of those described by AMIEVA, MOKRI et al. (2014) showing a cognitive decline up to 16 years before the diagnosis of dementia in highly educated individuals in the PAQUID cohort. This could indicate that subtle changes, related to AD brain lesions occurring up to 30 years before diagnosis (BATEMAN et al., 2012) would be recognized as psychiatric symptoms and treated as such. On argument in favor of this hypothesis is the prescription probability curve of antidementia drugs compared to that of the CN group. We see that the two curves diverge around 8 years before the diagnosis. This implies that the general practitioners detect subtle cognitive changes in some patients, years before they later decline to the point of AD dementia. This pre-AD diagnosis period of 5 to 10 years exactly matches the duration of the prodromal phase of the disease estimated recently in a large, multicohort study by VERMUNT et al. (2019). This means that it is in fact possible to diagnose AD earlier which would help in secondary prevention trials. Nowadays, the frequency of patients with early stage AD diagnosis in France is quite low for many reasons, including the low referral by general practitioners to memory clinic specialists (EPELBAUM, PAQUET et al., 2019).

4.4.2 Prediction

In our study, the simple algorithmic analysis of the combination of studied drug categories prescription yielded fair screening performances for further AD diagnosis in the 5 following years. Of note, we selected the model that was the most clinically pertinent, selecting both from the AD and from the stable MCI groups as patients. This finding has major public health implications as it opens new opportunities to screen for dementia in the elderly in a simple, implicit fashion and at no cost. Thus far, screening for dementia or identifying at-risk for dementia individuals relies on genetic (ESCOTT-PRICE et al., 2015), clinical (JOHNSON et al., 2014) or neuroradiological investigations (ARDEKANI et al., 2017; SAMPER-GONZALEZ et al., 2019; CHINCARINI et al., 2011). Integrating the screening process to routine practice has many advantages compared to these techniques which all require the active participation of patients and are costly. Our algorithm could seamlessly alert the general practitioner about the risk of further dementia which

would allow to enrich prevention trials in “at-risk” participants and result in decreases in recruitment cost (ANSART, EPELBAUM, GAGLIARDI, COLLIOT, DORMONT, DUBOIS et al., 2019).

4.4.3 Management practices

At the time of diagnosis, we evidenced a spectacular increase in psychotropic prescription in AD patients. Although this seems coherent for antedementia drugs, this is much more surprising for antipsychotics which use is advised against by French and European healthcare authorities since 2008 (ANKRI and VAN BROECKHOVEN 2013). This is probably due to the fact that the Cegedim aggregates data from the last 25 years and it will be a particularly useful tool to monitor this practice, which can be impacted by public health policies (DONEGAN et al., 2017), in the coming years.

The decrease in almost all drug categories prescription probably reflects the gradual changes induced by the autonomy loss over the course of AD. The general practitioners tend to simplify the therapeutic procedures as much as possible for these patients, especially in institutions (MASSOT MESQUIDA et al., 2019). The decrease in antedementia drugs probably relates to the limited magnitude of effect (BIRKS et GRIMLEY EVANS, 2015; KISHI et al., 2017) which can sometime be disappointing for patients and their care giver, and lead to treatment discontinuation.

In fact, most treatment categories display a decreasing slope of prescription after AD diagnosis which seems opposed to recent findings in a recent observational study of prescription changes following nursing home admission (ATRAMONT et al., 2018). However, our study does not indicate if patients were institutionalized or not which explains part of the discrepancy. One should also note that despite this gradual post-diagnosis prescription decrease, the frequency of psychotropic drugs remained higher in AD patients than in the two control groups as already described (RENOM-GUITERAS et al., 2018).

4.4.4 Strengths and weaknesses of the study

The use of a large sample of patients representative of the general population in France assessed with the same standardized electronic clinical records software, is among the main strengths of our study.

Another strength lies in the use of three groups rather than two. In most populational studies the model analyzes differences between one group with a condition and a control group (TZENG et al., 2018; PERERA et al., 2014; W.-Y. LIN et al., 2019). In AD research however, such a dichotomy does not consider the complexity of this affection. Prior to dementia, stages of preclinical and prodromal AD

(or MCI due to AD) have been described (DUBOIS, FELDMAN, JACOVA, DEKOSKY et al., 2007; DUBOIS, FELDMAN, JACOVA, HAMPEL et al., 2014; JACK, BENNETT et al., 2018). These stages can sometime be difficult to diagnose. Roughly 50% of patients with MCI have a genuine AD process (PETERSEN et al., 2013). Using a stable MCI control group allowed us to distinguish “chronic conditions affecting cognition” (such as lasting psychiatric conditions such as anxiety of recurring depression, learning disability, traumatic brain injuries) from neurodegenerative disorders leading to dementia. In the stable MCI group for instance psychotropic drugs are initially more frequently prescribed than in the AD group. However, this prescription frequency remains stable over time whereas that in the AD group gradually increases and exceeds it in the 10 to 5-year period before AD diagnosis. Selecting a CN group allowed us to evidence subtle differences with the AD group (notably concerning anti herpetic, anti-inflammatory and antirheumatic, glucose and tension lowering drugs) which might have otherwise remained obfuscated.

Finally, the long period of follow-up is particularly well suited for the study of such a chronic disease as AD spanning decades of life (VERMUNT et al., 2019).

As in all large scale, populational studies, the diagnosis of AD remains however based mostly on its classical, mostly clinical criteria and have not systematically been validated in expert memory clinics with the latest biomarkers. However, as in genome wide association studies, the relative lack of precision of data is well compensated by the large sample size which allows to draw general conclusions. Finally, the retrospective case control studies do not permit to draw causality inferences from their findings. For instance, as previously discussed, the over prescription of antidepressant in the AD group 15 years before diagnosis could be the cause or consequence (and maybe even both) of AD later in life. Only intervention studies and the longitudinal follow-up of patients (in the case of AD for decades) might be of value in informing on the directionality of the observed associations.

4.5 Conclusion

This large scale naturalistic observational study is informative on the prescription practices associated with AD diagnosis. Some of our findings can be interpreted as putative risk factors of the disease while others are more probably related to healthcare practices and recommendations. We also introduced the concept that healthcare monitoring over long periods of time could be used to screen for dementia. Such large standardized routinely sustained databases will certainly prove to be very valuable tools to develop and validate public health policies in the future.

Conclusion & Perspectives

Conclusions

We proposed several clinical decision support systems to automatically identify groups of at risk individuals based on different criteria. We first considered the methodological issues that the design of such systems entails, and identified best practices by conducting a review of studies which perform an automatic prediction of the future diagnosis of MCI subjects. We then take advantage of our findings to propose our own method for performing this prediction, and compared several methodological options in a simple framework. Thirdly, we proposed a method for selecting individuals at risk of being amyloid positive, in order to recruit subjects for clinical trials at a lower cost. These decision support systems were tested on clinical research cohort, which do not always reflect the clinical practice. In a last study, we therefore focused on electronic health records, and used treatment prescriptions to select individuals who are at risk of developing AD in the next 5 to 10 years. We summarize here our conclusions regarding each of these studies.

In a first study, we conducted a systematic and quantitative review of the methods which automatically predict the progression of mild cognitive impairment to Alzheimer's disease. We found that predictions based on MRI only performed significantly worse than others. These findings question the wide use of MRI in this field, and call for further exploration of cognitive assessments, which can be easily gathered and lead to a good performance. We identified several methodological issues, which pertain to the misuse of the test set during the training phase, or to the usability of the method in clinical practice. We propose guidelines to resolve these issues, and highlight the importance of following machine learning best practices. We show that short term predictions are not likely to perform better than predicting that all individuals stay stable over time, showing the importance of comparing the methods to this constant prediction. We also highlight a possible bias regarding the non publication of methods resulting in a low performance on a large data set.

In a second study, we proposed a method for automatically predicting the future diagnosis of MCI subjects, by first predicting their future cognitive scores. This approach gives a more complete view of how the patient is likely to evolve, which can be used for diagnosis but also for patient stratification or selection of a particular subgroup, and to tailor patient care at the individual level. This two-step prediction is also more interpretable for clinicians. By reducing the black-box effect, it is more likely to be used in clinical practice. Within this prediction framework we benchmarked a range of methodological options and assessed the performance on the prediction of the progression to AD at one year. We showed that using longitudinal information did not improve the results compared to using one visit only for prediction. Overall, using more complex features, which can be less available in clinical practice, did not lead to a better prediction than the one obtained using the simple framework.

An interesting perspective of this study would be to assess the performance of more complex methods regarding the use of imaging or longitudinal information. Deep learning methods for example, have been especially known for their good performance on image analysis on data sets of more than 50,000 samples (LAUZON, 2012; LI DENG, 2012). They could be used in our framework to automatically extract the most relevant features from MRI, although they usually give the best results when applied on a data set larger than the ADNI. In a similar manner, algorithms modeling the temporal changes of each individual using their full history could be used to improve the longitudinal prediction.

In a third study, we proposed a method for recruiting subjects for clinical trials such as to minimize recruitment costs. In this method, we first automatically select individuals with a higher risk of being amyloid positive, and then perform a PET scan on these individuals only to confirm their amyloid status. We tested our approach on three different cohorts and showed that using it to select individuals for clinical trials can lead to a 20% reduction in recruitment costs. We found that using cognition, socio-demographic information and Apoe4 leads to a lower recruitment cost than integrated MRI features or longitudinal data. We showed that the cohort selected using our method is representative and does not significantly differ from the cohort that would be selected by performing a PET scan to all possible individuals, and that it generalizes well when applied to new subject.

A limitation of the study is the data set size. Data set size can greatly impact the performance, and testing our method on a larger cohort, coming from a Phase 3 clinical trials for example could lead to even better results. We also tested only basic MRI features, and methods taking advantage of the full MRI to extract features

that are meaningful to amyloidosis prediction could lead to a better performance using MRI.

In a fourth study, we modeled the treatment patterns of AD, MCI and CN subjects from french medical records. We first studied the difference in treatment between these groups, and the changes in treatment occurring at the time of AD diagnosis, in order to identify risk factors and management practices in french health care. We showed that AD diagnosis resulted in a radical change in patient care. Differences in prescription between AD and MCI or CN patients can be observed up to 15 years before AD diagnosis, suggesting that the temporal horizon that is currently considered in clinical studies and trials is in fact too short. Studies spanning over at least a decade could give a better view of the long term changes in patients, and highlight changes in biomarkers that are not visible on a smaller time scale.

We then built a model to predict if a patient will develop AD in the coming 5 or 10 years, based on 6 months of treatment history. The adoption of such a system could help clinicians identify at-risk individuals, who could benefit from additional exams and a more rigorous monitoring. It could also constitute an interesting tool for selecting patients for clinical trials, by creating cohort with a higher proportion of individuals progressing to AD.

The main limitation of this study lies in its observational and retrospective nature. Conclusions can only be drawn regarding the correlation between events and not regarding their causality.

Perspectives

A large number of automatic methods have been proposed to diagnose Alzheimer's disease and to identify individuals at risk of progressing to AD. The machine learning community provides a range of performance measures that can be used to evaluate these methods using an objective metric. As a result, research on clinical decision support systems often aims at maximizing these performance measures. Although these performance metrics are important to consider, maximizing them is not the goal per se. When building such a decision support system, one should also consider how it can be meaningful to the clinical practice, and test it in conditions that best reflect its future use.

Proposed methods are widely tested on clinical research cohorts, which are easily available. However, a method trained on such a cohort cannot be expected to perform well in clinical practice, were the available features can be very different

and patients can have different characteristics. Clinical decision support systems would therefore gain at focusing on electronic health records (EHR), which are representative of the clinical practice and of the patients on which we aim at making predictions. Our work on the Cegedim data base is a good first step in this direction. We focused on treatment prescriptions, which are easily available, but other data types could contain additional information. The study of patient hospitalization, and the integration of biomarkers in the analysis could lead to a more refined identification of patients at risk of developing AD.

The wide adoption of such methods requires the evaluation of the cultural biases of medical practices. We have brought into light management practices identified in the Cegedim cohort, but these practices might be unique to French health care, and their generalization to the health care system of other countries is unclear. Before being ready to be used on a larger scale, decision support systems should also be tested in a prospective study. In a test framework, clinicians could receive an alert when an individual at risk of developing AD is identified, so that patient care can be tailored. The system could also suggest additional tests, such as cognitive questionnaires, in order to refine the prediction. The deployment of the system for testing could allow to evaluate its impact on early AD diagnosis and on the implementation and evaluation of new therapeutic strategies or prevention measures.

Lastly, our work has focused on the context of Alzheimer's disease, which is the most common neuro-degenerative disease. However, when a patient shows cognitive symptoms, practitioners are interested in knowing which disease the patient is likely to develop, and rarely focus on one condition in particular. The generalization of our work to differential diagnosis could therefore represent an interesting perspective.

Annexe A

Supplementary materials for the systematic and quantitative review

A.1 Query

The full query was :

```
TITLE-ABS-KEY ("alzheimer's" OR alzheimer OR ad) AND TITLE-ABS-KEY
  ("Mild Cognitive Impairment" OR "MCI") AND TITLE-ABS-KEY ((
  predicting OR prediction OR predictive) AND (conversion OR
  decline OR progression OR onset) OR prognosis) AND TITLE-ABS-KEY
  (accuracy OR roc OR auc OR specificity OR sensitivity) AND (
  TITLE-ABS-KEY ("Deep learning" OR "neural network" OR "neural
  networks" OR "convolutional network" OR "convolutional networks"
  OR "bayesian network" OR "bayesian networks") OR TITLE-ABS-KEY
  ("Matrix completion" OR "Support vector machine" OR "linear
  mixed-effect" OR "logistic regression" OR "Random Forest" OR "
  kernel classifier" OR "kernel" OR "decision tree" OR "decision
  trees" OR "least-squares") OR TITLE-ABS-KEY ("Machine learning"
  OR "pattern recognition" OR "pattern classification" OR "
  classifier" OR "algorithm" OR "classification"))
```

A.2 Selection process diagram

The process used to select the articles included in the review is shown in [Figure A.1](#).

A.3 Reported items

For each article, the following elements were reported :

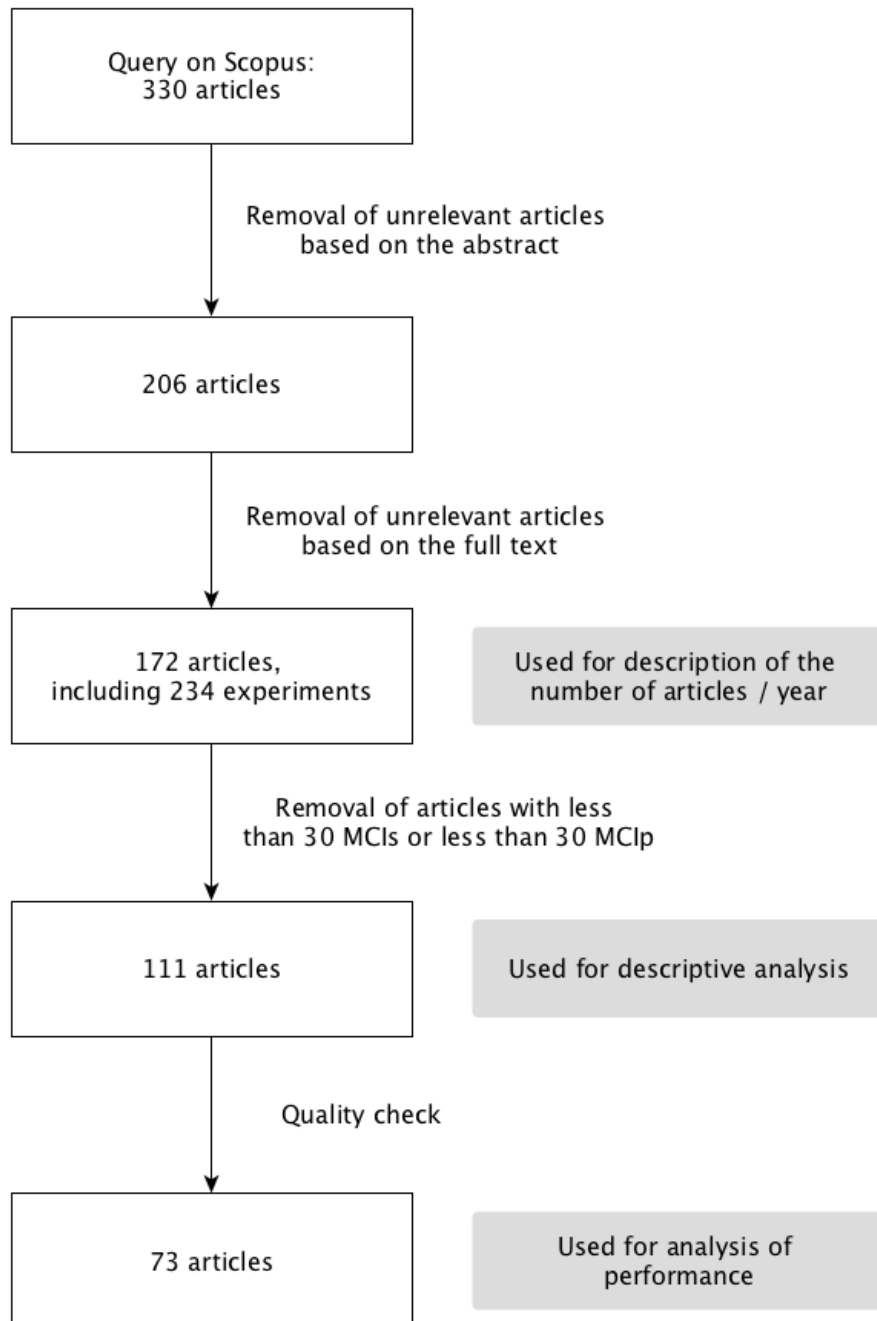


FIGURE A.1 – Diagram representing who the articles were selected

- number of MCI subjects progressing to AD;
- number of stable MCI subjects;
- time to prediction;
- used cohorts;
- use of socio-demographic features (yes/no);
- use of APOE (yes/no);
- use of general cognitive features (yes/no);
- use of domain-targeted cognitive features (yes/no);
- use of new, home-made cognitive features (yes/no);
- use of voxel based features from T1 MRI (yes/no);
- use of regions of interest on the whole brain, from T1 MRI (yes/no);
- use of selected regions of interest from T1 MRI (yes/no);
- use of white matter hyper-intensities (yes/no);
- use of PET FDG features (yes/no);
- use of PET amyloid features (yes/no);
- use of PET tau features (yes/no);
- use of CSF features (yes/no);
- use of amyloid status (yes/no);
- use of DTI features (yes/no);
- use of functional MRI features (yes/no);
- use of EEG or MEG features (yes/no);
- use of other features (yes/no, precision given as a free note);
- use of longitudinal features (yes/no);
- is feature selection performed (yes/no);
- used algorithm (categories defined below);
- validation method (categories defined bellow);
- feature selection performed on the whole data set (yes/no/unclear);
- feature embedding performed on the whole data set (yes/no/unclear);
- selection of the input visit of the test subjects using their date of progression to AD (yes/no);
- other data leakage (use of the test set to make decisions) (yes/no/unclear);
- other issue (yes/no)
- AUC value;
- accuracy value;
- balanced accuracy value;
- sensitivity value;
- specificity value;

Free notes describing the issues, or important points that did not fit in the previous list, were added.

The possible algorithm categories were added by the readers and aggregated. The final list was : bayesian algorithms, classification by clinicians, gaussian process, linear discriminant analysis (LDA), low rank matrix completion (LRMC), linear regression, logistic regression, manifold learning, multiple kernel learning, neural network, orthogonal partial least square (OPLS), random forest, regularized logistic regression, support vector machine, survival analysis, use of a threshold and others (including home-made algorithms).

The same process was used to create the cross-validation category list, composed of : 10-fold, k-fold, repeated k-fold, leave one out, out of the bag, single split, repeated single split, validation on an independent cohort, validation on different groups (when the algorithm is trained on separating AD and CN subjects, and tested on predicting the progression of MCI subjects), none, not described (when the use of cross-validation is mentioned but the used validation method is not described) and not needed (for thresholding with a manually chosen threshold for example).

A.4 Journals and conference proceedings

Table A.1 shows the journals and conference proceedings in which more than one included article has been published, and the associated number of articles.

A.5 Information table

A table containing all the articles included in the review and all the reported values can be found on <https://gitlab.com/icm-institute/aramislab/mci-progression-review>. The issues identified in each articles were removed from this open-access table, to avoid negatively pointing at these studies. They can be made available if requested to the corresponding author.

Journal or conference proceedings	Number of included articles
Journal of Alzheimer's Disease	12
NeuroImage	11
Lecture Notes in Computer Science	7
PLoS ONE	9
Neurobiology of Aging	6
Neurology	3
Brain Topography	3
Current Alzheimer Research	3
Medical Image Analysis	3
Frontiers in Aging Neuroscience	3
Scientific Reports	2
Frontiers in Neuroscience	2
IEEE Journal of Biomedical and Health Informatics	2
IEEE Transactions on Biomedical Engineering	2
NeuroImage : Clinical	2
Journal of Neuroscience Methods	2

TABLE A.1 – Number of included articles published in each journal or conference proceedings. Only the journals with more than one included article are shown here. The articles taken into account are the one considered for analysis, and that use a large enough data set.

Annexe B

Supplementary materials for amyloidosis prediction

B.1 Computing R and S from the PPV and NPR

The number of False Positives (FP) can be computed from the Positive Predicted Value (PPV) and the number of True Positives (TP) as such :

$$PPV = \frac{TP}{FP + TP}$$

$$TP = PPV * FP + PPV * TP$$

$$FP = \frac{1 - PPV}{PPV} TP \quad (B.1)$$

In a similar manner, the number of False Negatives (FN) can be computed from the Negative Predicted Value (NPV) and the number of True Negatives (TN) :

$$TN = \frac{NPV}{1 - NPV} * FN \quad (B.2)$$

We know that, NP being the number of positive subjects in the test set,

$$FN = NP - TP \quad (B.3)$$

And, N being the total number of subjects in the test set :

$$FP + FN + TP + TN = N \quad (B.4)$$

Using equations 3 to 6, we can deduce

$$TP = (N(1 - NPV) - NP) \frac{PPV}{1 - NPV - PPV}$$

And S and R and be computed using equations 1 and 2.

B.2 Difference of age in the 3 cohorts

	Age average for $A\beta^-$ individuals (std)	Age average for $A\beta^+$ individuals (std)	p-value
INSIGHT	75.7 (3.5)	76.8 (3.4)	0.01
ADNI-CN	74.4 (6.5)	76.2 (6.1)	0.005
ADNI-MCI	72.0 (8.5)	74.7 (6.9)	< 0.001

TABLE B.1 – Age comparison between $A\beta^-$ and $A\beta^+$ individuals for the different cohorts. std = standard deviation.

B.3 Algorithm pseudo-code

Algorithm 1 Pseudocode of the method**Input :** x and y **Output :** $probs$: probability of each subject to be $A\beta+$; auc : the obtained AUC; min_cost : minimal cost for recruiting the subjects; $optimal_threshold$: probability threshold for which the minimal cost is obtained**for** $i = 1$ to 50 **do**

▷ Randomly split into training and test set, with 30% in test set

 $x_train, y_train, x_test, y_test \leftarrow split(x, y, 0.3)$

▷ Hyper-parameter tuning using the AUC

 $splits_x, splits_y \leftarrow split_in_5(x_train, y_train)$ **for** $num_fold = 1$ to 5 **do**

▷ Get the corresponding folds for training and testing

 $x_test_fold, y_test_fold \leftarrow splits_x[num_fold], splits_y[num_fold]$ $x_train_fold \leftarrow all_folds_except_i(splits_x, num_fold)$ $y_train_fold \leftarrow all_folds_except_i(splits_y, num_fold)$ **for** $i_size = 1$ to number_leaf_sizes **do****for** $i_cycles = 1$ to number_num_cycles **do**

▷ Train and predict with the selected parameters

 $rf \leftarrow fit_rf(x_train_fold, y_train_fold, leaf_sizes[i_size], num_cycles[i_cycles])$ $probs \leftarrow get_rf_score(rf, x_test_fold)$

▷ Compute the corresponding AUC

 $auc \leftarrow get_auc(probs, y_test_fold)$ $aucs_table.insert(auc)$ **end for****end for****end for**

▷ Average the AUC for each parameters over all folds

 $mean_aucs \leftarrow average_over_folds(aucs_table)$

▷ Select the parameter values corresponding to the best AUC

 $i_best_size, i_best_num_cycles \leftarrow argmax(mean_aucs)$ $leaf_size, num_cycle \leftarrow leaf_sizes[i_best_size], num_cycles[i_best_num_cycles]$

▷ Train and apply the model with the selected hyper-parameters

 $rf \leftarrow train_rf(x_train, y_train, leaf_size, num_cycles)$ $probs \leftarrow get_rf_score(rf, x_test)$ $auc \leftarrow get_auc(probs, y_test)$

▷ Get the threshold for minimal cost

 $sen_table, spe_table, thresholds_table \leftarrow get_all_sensitivities_specificities(probs, y_test)$ $costs_table \leftarrow all_possible_costs(sen_table, spe_table)$ $min_cost \leftarrow min(costs_table)$ $i_min \leftarrow argmin(costs_table)$ $optimal_threshold \leftarrow thresholds_table[i_min]$ **end for**

Annexe C

Supplementary materials for the study of treatment prescriptions

C.1 Statistical analysis

C.1.1 Model description

Let $Y_{i,j}$ be binary variable with value 1 when the subject i has a prescription at the j^{th} time point, and 0 otherwise. $Y_{i,j}$ follows a Bernoulli distribution with $P(Y_{i,j} = 1) = \mu_{i,j}$. $\mu_{i,j}$ is modeled as :

$$\log\left(\frac{\mu_{i,j}}{1 - \mu_{i,j}}\right) = \beta_1 + \beta_2 t_{i,j} + \beta_3 AD_i + \beta_4 AD_i t_{i,j} + \beta_5 AD_i (t_{i,j})_+ + \beta_6 AD_i (t_{i,j})_+ t_{i,j} + b_i$$

with $AD_i = 1$ for AD patients and 0 for other patients, and $(t_{i,j})_+ = 1$ when $t_{i,j} > 0$ and 0 otherwise.

The model is composed of 4 main parts. The first part, $\beta_1 + \beta_2 t_{i,j}$, corresponds to a linear regression common to all subjects fitted by the model. The second part, $\beta_3 AD_i + \beta_4 AD_i t_{i,j}$, corresponds to the difference between the AD subjects, for which $AD_i = 1$ and the subjects of the other group fitted in the model (MCI or CN), for which $AD_i = 0$, in terms of intercept and slope. The third part, $\beta_5 AD_i (t_{i,j})_+ + \beta_6 AD_i (t_{i,j})_+ t_{i,j}$ corresponds to the change in response for the AD subjects after their diagnosis, when $t_{i,j} > 0$ and $t_{i,j} = 1$, in terms of intercept and slope. Lastly, b_i corresponds to the random intercept for subject i .

This model therefore accounts for the difference between groups and the change after diagnosis for the AD patients. Two models are fitted : one for the comparison between the MCI and AD patients, and one for the comparison between CN and AD patients.

C.1.2 Coefficient interpretation

For all the following calculations, we note $P_{AD}(t)$ the probability of receiving a given treatment at time t for a subject of the AD group, and $P_{\overline{AD}}(t)$ the same probability for a subject of another group. In a similar way, we note $o_{AD}(t)$ the odds of receiving the treatment at time t for a subject of the AD group, and $o_{\overline{AD}}(t)$ for a subject of another group. For a given subject i of any group,

$$o(t) = \frac{P(t)}{1 - P(t)} = e^{\beta_1 + \beta_2 t + \beta_3 AD_i + \beta_4 AD_i t + \beta_5 AD_i(t)_+ + \beta_6 AD_i(t)_+ t + b_i} \quad (C.1)$$

C.1.3 Intercept of the non-AD group

$$o_{\overline{AD}}(0) = \frac{P_{\overline{AD}}(0)}{1 - P_{\overline{AD}}(0)} = e^{\beta_1 + b_i} \quad (C.2)$$

Estimation of the expectation :

$$E\left(e^{\beta_1 + b_i}\right) = e^{\beta_1} E\left(e^{b_i}\right) = e^{\beta_1} \quad (C.3)$$

For standard deviation estimation :

$$\beta_1 \sim \mathcal{N}\left(\mu_{\beta_1}, \sigma_{\beta_1}\right) \quad (C.4)$$

$$b_i \sim \mathcal{N}\left(\mu_{b_i}, \sigma_{b_i}\right) \quad (C.5)$$

$$e^{\beta_1 + b_i} \sim \mathcal{N}\left(e^{\mu_{\beta_1} + \mu_{b_i} + \frac{\sigma_{\beta_1}^2 + \sigma_{b_i}^2}{2}}, \left(e^{\sigma_{\beta_1}^2 + \sigma_{b_i}^2} - 1\right) e^{2\mu_{\beta_1} + 2\mu_{b_i} + \sigma_{\beta_1}^2 + \sigma_{b_i}^2}\right) \quad (C.6)$$

C.1.4 Slope of the non-AD group

$$\frac{o_{\overline{AD}}(t+1)}{o_{\overline{AD}}(t)} = \frac{e^{\beta_1 + \beta_2(t+1) + b_i}}{e^{\beta_1 + \beta_2 t + b_i}} = e^{\beta_2} \quad (C.7)$$

C.1.5 Intercept change for the AD group

In order to only consider the changes due to belonging to the AD group without the effect of the AD diagnosis, we take $t < 0$

$$\frac{o_{AD}(0)}{O_{AD}(0)} = \frac{e^{\beta_1 + \beta_3 + b_i}}{e^{\beta_1 + b_i}} = e^{\beta_3} \quad (C.8)$$

C.1.6 Slope change for the AD group

Slope for the AD group, at $t < 0$:

$$s_{AD} = \frac{o_{AD}(t+1)}{o_{AD}(t)} = \frac{e^{\beta_1 + \beta_2(t+1) + \beta_3 + \beta_4(t+1) + b_i}}{e^{\beta_1 + \beta_2 t + \beta_3 + \beta_4 t + b_i}} = e^{\beta_2 + \beta_4} \quad (\text{C.9})$$

Slope for the non-AD group, such as detailed in C.7 :

$$s_{AD} = e^{\beta_2} \quad (\text{C.10})$$

Hence,

$$\frac{S_{AD}}{s_{AD}} = \frac{e^{\beta_2 + \beta_4}}{e^{\beta_2}} = e^{\beta_4} \quad (\text{C.11})$$

C.1.7 Impact of diagnosis on the intercept

We aim to measure the change of intercept in the AD group after AD diagnosis :

$$\frac{o_{AD}(0^+)}{o_{AD}(0^-)} = \frac{e^{\beta_1 + \beta_3 + \beta_5}}{e^{\beta_1 + \beta_3}} = e^{\beta_5} \quad (\text{C.12})$$

C.1.8 Impact of diagnosis on the slope

We note s_{AD}^+ the slope for a subject of the AD group and for a time $t > 0$:

$$s_{AD}^+ = \frac{o_{AD}(t+1)}{o_{AD}(t)} = \frac{e^{\beta_1 + \beta_2(t+1) + \beta_3 + \beta_4(t+1) + \beta_5 + \beta_6(t+1)}}{e^{\beta_1 + \beta_2 t + \beta_3 + \beta_4 t + \beta_5 + \beta_6 t}} = e^{\beta_2 + \beta_4 + \beta_6} \quad (\text{C.13})$$

We note s_{AD}^- the slope for a subject of the AD group and for a time $t < -1$:

$$s_{AD}^- = \frac{o_{AD}(t+1)}{o_{AD}(t)} = \frac{e^{\beta_1 + \beta_2(t+1) + \beta_3 + \beta_4(t+1)}}{e^{\beta_1 + \beta_2 t + \beta_3 + \beta_4 t}} = e^{\beta_2 + \beta_4} \quad (\text{C.14})$$

Hence,

$$\frac{s_{AD}^+}{s_{AD}^-} = \frac{e^{\beta_2 + \beta_4 + \beta_6}}{e^{\beta_2 + \beta_4}} = e^{\beta_6} \quad (\text{C.15})$$

C.2 Predictive model

C.2.1 Performance measures

Definition of the performance measures, with TP = number of True Positives, FP = number of False Positive, TN = number of True Negatives, FN = number of

False Negatives, and J = Youden's J statistic.

$$\text{sensitivity} = \frac{TP}{TP + FN} \quad (\text{C.16})$$

$$\text{specificity} = \frac{TN}{TN + FP} \quad (\text{C.17})$$

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (\text{C.18})$$

$$\text{balanced accuracy} = \frac{\text{sensitivity} + \text{specificity}}{2} \quad (\text{C.19})$$

$$J = \text{sensitivity} + \text{specificity} - 1 \quad (\text{C.20})$$

C.2.2 Results optimized for screening

In the results shown in section 4.3, we used Youden's method to choose the point on the ROC curve, hence maximizing Youden's J statistic (J). Table C.1 shows the results obtained by maximizing the sensitivity, for a specificity of at least 80%.

	AUC	acc	bacc	sen	spe
AD and MCI groups (5 years)	70.8 (0.61)	65.2 (0.56)	65.3 (0.56)	50.5 (1.1)	80 (0.04)
AD and CN groups (5 years)	70.5 (0.6)	70.8 (0.32)	64.6 (0.55)	49.1 (1.1)	80 (0.035)
AD group (5 years)	69.2 (0.85)	58.2 (1.2)	64.3 (0.89)	48.5 (1.8)	80.1 (0.045)
AD group (10 years)	75.6 (1.1)	62.1 (2.6)	70.4 (1.4)	60.6 (2.8)	80.1 (0.1)

TABLE C.1 – CPerformance of the prediction of the presence of an AD diagnosis 5 or 10 years after a random visit for different groups of subjects. AUC = area under the receiver operating characteristic curve; acc = accuracy; bacc = balanced accuracy; sen = sensitivity; spe = specificity.

Annexe D

Articles included in the review

Complete list of articles included in the review of automatic prediction of the progression of Mild Cognitive Impairment:

- Aguilar, C., J.-S. Muehlboeck, et al. (2014). “Application of a MRI based index to longitudinal atrophy change in Alzheimer disease, mild cognitive impairment and healthy older individuals in the AddNeuroMed cohort”. In: *Frontiers in Aging Neuroscience* 6.JUL. DOI: [10.3389/fnagi.2014.00145](https://doi.org/10.3389/fnagi.2014.00145).
- Aguilar, C., E. Westman, et al. (2013). “Different multivariate techniques for automated classification of MRI data in Alzheimer’s disease and mild cognitive impairment”. en. In: *Psychiatry Research: Neuroimaging* 212.2, pp. 89–98. DOI: [10.1016/j.psychresns.2012.11.005](https://doi.org/10.1016/j.psychresns.2012.11.005).
- Arbizu, J. et al. (2013). “Automated analysis of FDG PET as a tool for single-subject probabilistic prediction and detection of Alzheimer’s disease dementia”. In: *European Journal of Nuclear Medicine and Molecular Imaging* 40.9, pp. 1394–1405. DOI: [10.1007/s00259-013-2458-z](https://doi.org/10.1007/s00259-013-2458-z).
- Arco, J. et al. (2016). “Short-term prediction of MCI to AD conversion based on longitudinal MRI analysis and neuropsychological tests”. In: *Smart Innovation, Systems and Technologies* 45, pp. 385–394. DOI: [10.1007/978-3-319-23024-5_35](https://doi.org/10.1007/978-3-319-23024-5_35).
- Ardekani, B. A. et al. (2017). “Prediction of Incipient Alzheimer’s Disease Dementia in Patients with Mild Cognitive Impairment”. en. In: *Journal of Alzheimer’s Disease* 55.1, pp. 269–281. DOI: [10.3233/JAD-160594](https://doi.org/10.3233/JAD-160594).
- Ayub, A. et al. (2017). “A novel approach for the prediction of conversion from mild cognitive impairment to Alzheimer’s disease using MRI images”. In: *Advances in Electrical and Computer Engineering* 17.2, pp. 113–122. DOI: [10.4316/AECE.2017.02015](https://doi.org/10.4316/AECE.2017.02015).
- Beer, J. et al. (2018). “Does a cognitive stress test predict progression from mild cognitive impairment to dementia equally well in clinical versus population-based settings?” In: *International Psychogeriatrics* 30.10, pp. 1435–1445. DOI: [10.1017/S1041610217002666](https://doi.org/10.1017/S1041610217002666).

- Beheshti, I., N. Maikusa, et al. (2017). “Classification of Alzheimer’s disease and prediction of mild cognitive impairment conversion using histogram-based analysis of patient-specific anatomical brain connectivity networks”. In: *Journal of Alzheimer’s Disease* 60.1, pp. 295–304. DOI: [10.3233/JAD-161080](https://doi.org/10.3233/JAD-161080).
- Beheshti, I., H. Demirel, and H. Matsuda (2017). “Classification of Alzheimer’s disease and prediction of mild cognitive impairment-to-Alzheimer’s conversion from structural magnetic resonance imaging using feature ranking and a genetic algorithm”. en. In: *Computers in Biology and Medicine* 83, pp. 109–119. DOI: [10.1016/j.compbiomed.2017.02.011](https://doi.org/10.1016/j.compbiomed.2017.02.011).
- Belleville, S. et al. (2014). “Predicting decline in mild cognitive impairment: A prospective cognitive study”. In: *Neuropsychology* 28.4, pp. 643–652. DOI: [10.1037/neu0000063](https://doi.org/10.1037/neu0000063).
- Borroni, B. et al. (2006). “Combined 99mTc-ECD SPECT and neuropsychological studies in MCI for the assessment of conversion to AD”. In: *Neurobiology of Aging* 27.1, pp. 24–31. DOI: [10.1016/j.neurobiolaging.2004.12.010](https://doi.org/10.1016/j.neurobiolaging.2004.12.010).
- Buratti, L. et al. (2015). “Markers for the Risk of Progression from Mild Cognitive Impairment to Alzheimer’s Disease”. In: *Journal of Alzheimer’s Disease* 45.3, pp. 883–890. DOI: [10.3233/JAD-143135](https://doi.org/10.3233/JAD-143135).
- Buscema, M. et al. (2010). “The I.F.A.S.T. model allows the prediction of conversion to alzheimer disease in patients with mild cognitive impairment with high degree of accuracy”. In: *Current Alzheimer Research* 7.2, pp. 173–187. DOI: [10.2174/156720510790691137](https://doi.org/10.2174/156720510790691137).
- Cabranes, J. et al. (2004). “Relevance of functional neuroimaging in the progression of mild cognitive impairment”. In: *Neurological Research* 26.5, pp. 496–501. DOI: [10.1179/016164104225016155](https://doi.org/10.1179/016164104225016155).
- Callahan, B. et al. (2015). “Predicting Alzheimer’s disease development: A comparison of cognitive criteria and associated neuroimaging biomarkers”. In: *Alzheimer’s Research and Therapy* 7.1. DOI: [10.1186/s13195-015-0152-z](https://doi.org/10.1186/s13195-015-0152-z).
- Caminiti, S. et al. (2018). “FDG-PET and CSF biomarker accuracy in prediction of conversion to different dementias in a large multicentre MCI cohort”. In: *NeuroImage: Clinical* 18, pp. 167–177. DOI: [10.1016/j.nicl.2018.01.019](https://doi.org/10.1016/j.nicl.2018.01.019).
- Chen, R. et al. (2012). “Prediction of conversion from mild cognitive impairment to Alzheimer disease based on Bayesian data mining with ensemble learning”. In: *Neuroradiology Journal* 25.1, pp. 5–16. DOI: [10.1177/197140091202500101](https://doi.org/10.1177/197140091202500101).
- Chen, X. et al. (2016). “Potential clinical value of multiparametric pet in the prediction of Alzheimer’s disease progression”. In: *PLoS ONE* 11.5. DOI: [10.1371/journal.pone.0154406](https://doi.org/10.1371/journal.pone.0154406).

- Cheng, B., M. Liu, H.-I. Suk, et al. (2015). “Multimodal manifold-regularized transfer learning for MCI conversion prediction”. In: *Brain Imaging and Behavior* 9.4, pp. 913–926. DOI: [10.1007/s11682-015-9356-x](https://doi.org/10.1007/s11682-015-9356-x).
- Cheng, B., M. Liu, D. Zhang, et al. (2015). “Domain Transfer Learning for MCI Conversion Prediction”. In: *IEEE Transactions on Biomedical Engineering* 62.7, pp. 1805–1817. DOI: [10.1109/TBME.2015.2404809](https://doi.org/10.1109/TBME.2015.2404809).
- Chincarini, A. et al. (2011). “Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer’s disease”. In: *NeuroImage* 58.2, pp. 469–480. DOI: [10.1016/j.neuroimage.2011.05.083](https://doi.org/10.1016/j.neuroimage.2011.05.083).
- Cho, Y. et al. (2012). “Individual subject classification for Alzheimer’s disease based on incremental learning using a spatial frequency representation of cortical thickness data”. In: *NeuroImage* 59.3, pp. 2217–2230. DOI: [10.1016/j.neuroimage.2011.09.085](https://doi.org/10.1016/j.neuroimage.2011.09.085).
- Choi, H. and K. Jin (2018). “Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging”. In: *Behavioural Brain Research* 344, pp. 103–109. DOI: [10.1016/j.bbr.2018.02.017](https://doi.org/10.1016/j.bbr.2018.02.017).
- Choo, I. et al. (2013). “Combination of 18F-FDG PET and cerebrospinal fluid biomarkers as a better predictor of the progression to Alzheimer’s disease in mild cognitive impairment patients”. In: *Journal of Alzheimer’s Disease* 33.4, pp. 929–939. DOI: [10.3233/JAD-2012-121489](https://doi.org/10.3233/JAD-2012-121489).
- Çitak-Er, F. et al. (2017). “A novel convolutional neural network model based on voxel-based morphometry of imaging data in predicting the prognosis of patients with mild cognitive impairment”. In: *Journal of Neurological Sciences* 34.1, pp. 52–69.
- Clark, D. (2011). “Residual vectors for Alzheimer disease diagnosis and prognostication”. In: *Brain and Behavior* 1.2, pp. 142–152. DOI: [10.1002/brb3.19](https://doi.org/10.1002/brb3.19).
- Clark, D., P. Kapur, et al. (2014). “Latent Information in Fluency Lists Predicts Functional Decline in Persons at Risk for Alzheimer Disease”. In: *Cortex; a journal devoted to the study of the nervous system and behavior* 55, pp. 202–218. DOI: [10.1016/j.cortex.2013.12.013](https://doi.org/10.1016/j.cortex.2013.12.013).
- Clark, D., P. McLaughlin, et al. (2016). “Novel verbal fluency scores and structural brain imaging for prediction of cognitive outcome in mild cognitive impairment”. In: *Alzheimer’s and Dementia: Diagnosis, Assessment and Disease Monitoring* 2, pp. 113–122. DOI: [10.1016/j.dadm.2016.02.001](https://doi.org/10.1016/j.dadm.2016.02.001).
- Collij, L. et al. (2016). “Application of machine learning to arterial spin labeling in mild cognitive impairment and Alzheimer disease”. In: *Radiology* 281.3, pp. 865–875. DOI: [10.1148/radiol.2016152703](https://doi.org/10.1148/radiol.2016152703).

- Convit, A. et al. (2000). “Atrophy of the medial occipitotemporal, inferior, and middle temporal gyri in non-demented elderly predict decline to Alzheimer’s disease”. In: *Neurobiology of Aging* 21.1, pp. 19–26. DOI: [10.1016/S0197-4580\(99\)00107-4](https://doi.org/10.1016/S0197-4580(99)00107-4).
- Costafreda, S. et al. (2011). “Automated hippocampal shape analysis predicts the onset of dementia in mild cognitive impairment”. In: *NeuroImage* 56.1, pp. 212–219. DOI: [10.1016/j.neuroimage.2011.01.050](https://doi.org/10.1016/j.neuroimage.2011.01.050).
- Cui, Y. et al. (2011). “Identification of Conversion from Mild Cognitive Impairment to Alzheimer’s Disease Using Multivariate Predictors”. In: *PLOS ONE* 6.7, e21896. DOI: [10.1371/journal.pone.0021896](https://doi.org/10.1371/journal.pone.0021896).
- Cuingnet, R. et al. (2011). “Automatic classification of patients with Alzheimer’s disease from structural MRI: A comparison of ten methods using the ADNI database”. In: *NeuroImage* 56.2, pp. 766–781. DOI: [10.1016/j.neuroimage.2010.06.013](https://doi.org/10.1016/j.neuroimage.2010.06.013).
- Davatzikos, C. et al. (2011). “Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification”. In: *Neurobiology of Aging* 32.12, 2322.e19–2322.e27. DOI: [10.1016/j.neurobiolaging.2010.05.023](https://doi.org/10.1016/j.neurobiolaging.2010.05.023).
- Devanand, D., X. Liu, et al. (2008). “Combining Early Markers Strongly Predicts Conversion from Mild Cognitive Impairment to Alzheimer’s Disease”. In: *Biological Psychiatry* 64.10, pp. 871–879. DOI: [10.1016/j.biopsych.2008.06.020](https://doi.org/10.1016/j.biopsych.2008.06.020).
- Devanand, D., G. Pradhaban, et al. (2007). “Hippocampal and entorhinal atrophy in mild cognitive impairment: Prediction of Alzheimer disease”. In: *Neurology* 68.11, pp. 828–836. DOI: [10.1212/01.wnl.0000256697.20968.d7](https://doi.org/10.1212/01.wnl.0000256697.20968.d7).
- Devanand, D., R. Van Heertum, et al. (2010). “^{99m}Tc hexamethyl-propylene-aminoxime single-photon emission computed tomography prediction of conversion from mild cognitive impairment to Alzheimer disease”. In: *American Journal of Geriatric Psychiatry* 18.11, pp. 959–972. DOI: [10.1097/JGP.0b013e3181ec8696](https://doi.org/10.1097/JGP.0b013e3181ec8696).
- Dierckx, E. et al. (2009). “Verbal cued recall as a predictor of conversion to Alzheimer’s disease in Mild Cognitive Impairment”. In: *International Journal of Geriatric Psychiatry* 24.10, pp. 1094–1100. DOI: [10.1002/gps.2228](https://doi.org/10.1002/gps.2228).
- Doyle, O. et al. (2014). “Predicting progression of Alzheimer’s disease using ordinal regression”. In: *PLoS ONE* 9.8. DOI: [10.1371/journal.pone.0105542](https://doi.org/10.1371/journal.pone.0105542).
- Drzezga, A. et al. (2005). “Prediction of individual clinical outcome in MCI by means of genetic assessment and ¹⁸F-FDG PET”. In: *Journal of Nuclear Medicine* 46.10, pp. 1625–1632.
- Duchesne, S. et al. (2010). “Amnesic MCI future clinical status prediction using baseline MRI features”. In: *Neurobiology of Aging* 31.9, pp. 1606–1617. DOI: [10.1016/j.neurobiolaging.2008.09.003](https://doi.org/10.1016/j.neurobiolaging.2008.09.003).
- Dukart, J., F. Sambataro, and A. Bertolino (2015). “Accurate prediction of conversion to Alzheimer’s disease using imaging, genetic, and neuropsychological biomarkers”.

- In: *Journal of Alzheimer's Disease* 49.4, pp. 1143–1159. DOI: [10.3233/JAD-150570](https://doi.org/10.3233/JAD-150570).
- Eskildsen, S., P. Coupé, V. Fonov, et al. (2015). “Structural imaging biomarkers of Alzheimer’s disease: Predicting disease progression”. In: *Neurobiology of Aging* 36.S1, S23–S31. DOI: [10.1016/j.neurobiolaging.2014.04.034](https://doi.org/10.1016/j.neurobiolaging.2014.04.034).
- Eskildsen, S., P. Coupé, D. García-Lorenzo, et al. (2013). “Prediction of Alzheimer’s disease in subjects with mild cognitive impairment from the ADNI cohort using patterns of cortical thinning”. In: *NeuroImage* 65, pp. 511–521. DOI: [10.1016/j.neuroimage.2012.09.058](https://doi.org/10.1016/j.neuroimage.2012.09.058).
- Ewers, M., K. Buerger, et al. (2007). “Multicenter assessment of CSF-phosphorylated tau for the prediction of conversion of MCI”. In: *Neurology* 69.24, pp. 2205–2212. DOI: [10.1212/01.wnl.0000286944.22262.ff](https://doi.org/10.1212/01.wnl.0000286944.22262.ff).
- Ewers, M., C. Walsh, et al. (2012). “Prediction of conversion from mild cognitive impairment to Alzheimer’s disease dementia based upon biomarkers and neuropsychological test performance”. In: *Neurobiology of Aging* 33.7, 1203–1214.e2. DOI: [10.1016/j.neurobiolaging.2010.10.019](https://doi.org/10.1016/j.neurobiolaging.2010.10.019).
- Falahati, F. et al. (2016). “The Effect of Age Correction on Multivariate Classification in Alzheimer’s Disease, with a Focus on the Characteristics of Incorrectly and Correctly Classified Subjects”. In: *Brain Topography* 29.2, pp. 296–307. DOI: [10.1007/s10548-015-0455-1](https://doi.org/10.1007/s10548-015-0455-1).
- Fayed, N. et al. (2017). “Magnetic resonance spectroscopy and brain volumetry in mild cognitive impairment. A prospective study”. In: *Magnetic Resonance Imaging* 38, pp. 27–32. DOI: [10.1016/j.mri.2016.12.010](https://doi.org/10.1016/j.mri.2016.12.010).
- Ferreira, F. et al. (2017). “Improving prognostic prediction from mild cognitive impairment to Alzheimer’s disease using genetic algorithms”. In: *Advances in Intelligent Systems and Computing* 616, pp. 180–188. DOI: [10.1007/978-3-319-60816-7_22](https://doi.org/10.1007/978-3-319-60816-7_22).
- Fleisher, A. et al. (2008). “Volumetric MRI vs clinical predictors of Alzheimer disease in mild cognitive impairment”. In: *Neurology* 70.3, pp. 191–199. DOI: [10.1212/01.wnl.0000287091.57376.65](https://doi.org/10.1212/01.wnl.0000287091.57376.65).
- Fritzsche, K. et al. (2010). “Automated MR morphometry to predict Alzheimer’s disease in mild cognitive impairment”. In: *International Journal of Computer Assisted Radiology and Surgery* 5.6, pp. 623–632. DOI: [10.1007/s11548-010-0412-0](https://doi.org/10.1007/s11548-010-0412-0).
- Frölich, L. et al. (2017). “Incremental value of biomarker combinations to predict progression of mild cognitive impairment to Alzheimer’s dementia”. In: *Alzheimer’s Research and Therapy* 9.1. DOI: [10.1186/s13195-017-0301-7](https://doi.org/10.1186/s13195-017-0301-7).

- Gallagher, D. et al. (2010). “Detecting prodromal Alzheimer’s disease in mild cognitive impairment: Utility of the CAMCOG and other neuropsychological predictors”. In: *International Journal of Geriatric Psychiatry* 25.12, pp. 1280–1287. DOI: [10.1002/gps.2480](https://doi.org/10.1002/gps.2480).
- García-Herranz, S., M. Díaz-Mardomingo, and H. Peraita (2016). “Neuropsychological predictors of conversion to probable Alzheimer disease in elderly with mild cognitive impairment”. In: *Journal of Neuropsychology* 10.2, pp. 239–255. DOI: [10.1111/jnp.12067](https://doi.org/10.1111/jnp.12067).
- Gaser, C. et al. (2013). “BrainAGE in Mild Cognitive Impaired Patients: Predicting the Conversion to Alzheimer’s Disease”. In: *PLoS ONE* 8.6. DOI: [10.1371/journal.pone.0067346](https://doi.org/10.1371/journal.pone.0067346).
- Gavidia-Bovadilla, G. et al. (2017). “Early prediction of Alzheimer’s disease using null longitudinal model-based classifiers”. In: *PLoS ONE* 12.1. DOI: [10.1371/journal.pone.0168011](https://doi.org/10.1371/journal.pone.0168011).
- Geslani, D. et al. (2005). “Mild cognitive impairment: An operational definition and its conversion rate to Alzheimer’s disease”. In: *Dementia and Geriatric Cognitive Disorders* 19.5-6, pp. 383–389. DOI: [10.1159/000084709](https://doi.org/10.1159/000084709).
- Griffith, H. et al. (2006). “Amnesic mild cognitive impairment: Diagnostic outcomes and clinical prediction over a two-year time period”. In: *Journal of the International Neuropsychological Society* 12.2, pp. 166–175. DOI: [10.1017/S1355617706060267](https://doi.org/10.1017/S1355617706060267).
- Guo, S. et al. (2017). “Conversion discriminative analysis on mild cognitive impairment using multiple cortical features from MR images”. In: *Frontiers in Aging Neuroscience* 9.MAY. DOI: [10.3389/fnagi.2017.00146](https://doi.org/10.3389/fnagi.2017.00146).
- Hall, A. et al. (2015). “Predicting progression from cognitive impairment to alzheimer’s disease with the disease state index”. In: *Current Alzheimer Research* 12.1, pp. 69–79. DOI: [10.2174/1567205012666141218123829](https://doi.org/10.2174/1567205012666141218123829).
- Hinrichs, C. et al. (2011). “Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population”. In: *NeuroImage* 55.2, pp. 574–589. DOI: [10.1016/j.neuroimage.2010.10.081](https://doi.org/10.1016/j.neuroimage.2010.10.081).
- Hojjati, S., A. Ebrahimzadeh, A. Khazae, and A. Babajani-Feremi (2017). “Predicting conversion from MCI to AD using resting-state fMRI, graph theoretical approach and SVM”. In: *Journal of Neuroscience Methods* 282, pp. 69–80. DOI: [10.1016/j.jneumeth.2017.03.006](https://doi.org/10.1016/j.jneumeth.2017.03.006).
- Hojjati, S., A. Ebrahimzadeh, A. Khazae, A. Babajani-Feremi, and D. N. I. Alzheimer’s (2018). “Predicting conversion from MCI to AD by integrating rs-fMRI and structural MRI”. In: *Computers in Biology and Medicine* 102, pp. 30–39. DOI: [10.1016/j.combiomed.2018.09.004](https://doi.org/10.1016/j.combiomed.2018.09.004).

- Huang, C. et al. (2002). “Cingulate cortex hypoperfusion predicts Alzheimer’s disease in mild cognitive impairment”. In: *BMC Neurology* 2. DOI: [10.1186/1471-2377-2-9](https://doi.org/10.1186/1471-2377-2-9).
- Huang, M. et al. (2017). “Longitudinal measurement and hierarchical classification framework for the prediction of Alzheimer’s disease”. In: *Scientific Reports* 7, p. 39880. DOI: [10.1038/srep39880](https://doi.org/10.1038/srep39880).
- Johnson, P. et al. (2014). “Genetic algorithm with logistic regression for prediction of progression to Alzheimer’s disease”. In: *BMC Bioinformatics* 15. DOI: [10.1186/1471-2105-15-S16-S11](https://doi.org/10.1186/1471-2105-15-S16-S11).
- Junwei Ding and Qiu Huang (2017). “Prediction of MCI to AD conversion using Laplace Eigenmaps learned from FDG and MRI images of AD patients and healthy controls”. en. In: *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*. Chengdu, China: IEEE, pp. 660–664. ISBN: 978-1-5090-6238-6. DOI: [10.1109/ICIVC.2017.7984638](https://doi.org/10.1109/ICIVC.2017.7984638).
- Kauppi, K. et al. (2018). “Combining polygenic hazard score with volumetric MRI and cognitive measures improves prediction of progression from mild cognitive impairment to Alzheimer’s disease”. In: *Frontiers in Neuroscience* 12.APR. DOI: [10.3389/fnins.2018.00260](https://doi.org/10.3389/fnins.2018.00260).
- Kebets, V. et al. (2015). “Predicting Pure Amnesic Mild Cognitive Impairment Conversion to Alzheimer’s Disease Using Joint Modeling of Imaging and Clinical Data”. In: *Proceedings - 2015 International Workshop on Pattern Recognition in NeuroImaging, PRNI 2015*, pp. 85–88. DOI: [10.1109/PRNI.2015.23](https://doi.org/10.1109/PRNI.2015.23).
- Khan, W., C. Aguilar, et al. (2015). “A subset of cerebrospinal fluid proteins from a multi-analyte panel associated with brain atrophy, disease classification and prediction in Alzheimer’s disease”. In: *PLoS ONE* 10.8. DOI: [10.1371/journal.pone.0134368](https://doi.org/10.1371/journal.pone.0134368).
- Khan, W., E. Westman, et al. (2015). “Automated Hippocampal Subfield Measures as Predictors of Conversion from Mild Cognitive Impairment to Alzheimer’s Disease in Two Independent Cohorts”. In: *Brain Topography* 28.5, pp. 746–759. DOI: [10.1007/s10548-014-0415-1](https://doi.org/10.1007/s10548-014-0415-1).
- Kim, J. W. et al. (2017). “Clinical Dementia Rating Orientation Score as an Excellent Predictor of the Progression to Alzheimer’s Disease in Mild Cognitive Impairment”. en. In: *Psychiatry Investigation* 14.4, p. 420. DOI: [10.4306/pi.2017.14.4.420](https://doi.org/10.4306/pi.2017.14.4.420).
- Klöppel, S. et al. (2015). “Applying automated MR-based diagnostic methods to the memory clinic: A prospective study”. In: *Journal of Alzheimer’s Disease* 47.4, pp. 939–954. DOI: [10.3233/JAD-150334](https://doi.org/10.3233/JAD-150334).

- Koppara, A. et al. (2015). “The latent dementia phenotype δ is associated with cerebrospinal fluid biomarkers of Alzheimer’s disease and predicts conversion to dementia in subjects with mild cognitive impairment”. In: *Journal of Alzheimer’s Disease* 49.2, pp. 547–560. DOI: [10.3233/JAD-150257](https://doi.org/10.3233/JAD-150257).
- Korolev, I. O. et al. (2016). “Predicting Progression from Mild Cognitive Impairment to Alzheimer’s Dementia Using Clinical, MRI, and Plasma Biomarkers via Probabilistic Pattern Classification”. en. In: *PLOS ONE* 11.2, e0138866. DOI: [10.1371/journal.pone.0138866](https://doi.org/10.1371/journal.pone.0138866).
- Kruczyk, M. et al. (2012). “Monte Carlo feature selection and rule-based models to predict Alzheimer’s disease in mild cognitive impairment”. In: *Journal of Neural Transmission* 119.7, pp. 821–831. DOI: [10.1007/s00702-012-0812-0](https://doi.org/10.1007/s00702-012-0812-0).
- Lehallier, B. et al. (2016). “Combined plasma and cerebrospinal fluid signature for the prediction of midterm progression from mild cognitive impairment to Alzheimer disease”. In: *JAMA Neurology* 73.2, pp. 203–212. DOI: [10.1001/jamaneurol.2015.3135](https://doi.org/10.1001/jamaneurol.2015.3135).
- Lei, B., S. Chen, et al. (2016). “Discriminative learning for Alzheimer’s disease diagnosis via canonical correlation analysis and multimodal fusion”. In: *Frontiers in Aging Neuroscience* 8.MAY. DOI: [10.3389/fnagi.2016.00077](https://doi.org/10.3389/fnagi.2016.00077).
- Lei, B., P. Yang, et al. (2017). “Relational-Regularized Discriminative Sparse Learning for Alzheimer’s Disease Diagnosis”. en. In: *IEEE Transactions on Cybernetics* 47.4, pp. 1102–1113. DOI: [10.1109/TCYB.2016.2644718](https://doi.org/10.1109/TCYB.2016.2644718).
- Li, H. et al. (2014). “Hierarchical interactions model for predicting Mild Cognitive Impairment (MCI) to Alzheimer’s Disease (AD) conversion”. In: *PLoS ONE* 9.1. DOI: [10.1371/journal.pone.0082450](https://doi.org/10.1371/journal.pone.0082450).
- Li, Y., J. Jiang, et al. (2018). “Radiomics features as predictors to distinguish fast and slow progression of Mild Cognitive Impairment to Alzheimer’s disease”. In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*. Vol. 2018-July, pp. 127–130. DOI: [10.1109/EMBC.2018.8512273](https://doi.org/10.1109/EMBC.2018.8512273).
- Li, Y., Z. Yao, et al. (2018). “Indirect relation based individual metabolic network for identification of mild cognitive impairment”. In: *Journal of Neuroscience Methods* 309, pp. 188–198. DOI: [10.1016/j.jneumeth.2018.09.007](https://doi.org/10.1016/j.jneumeth.2018.09.007).
- Lin, W. et al. (2018). “Convolutional neural networks-based MRI image analysis for the Alzheimer’s disease prediction from mild cognitive impairment”. In: *Frontiers in Neuroscience* 12.NOV. DOI: [10.3389/fnins.2018.00777](https://doi.org/10.3389/fnins.2018.00777).
- Liu, H. et al. (2016). “A semi-mechanism approach based on MRI and proteomics for prediction of conversion from mild cognitive impairment to Alzheimer’s disease”. In: *Scientific Reports* 6. DOI: [10.1038/srep26712](https://doi.org/10.1038/srep26712).

- Liu, X. et al. (2013). “Locally linear embedding (LLE) for MRI based Alzheimer’s disease classification”. In: *NeuroImage* 83, pp. 148–157. DOI: [10.1016/j.neuroimage.2013.06.033](https://doi.org/10.1016/j.neuroimage.2013.06.033).
- Long, X. et al. (2017). “Prediction and classification of Alzheimer disease based on quantification of MRI deformation”. In: *PLoS ONE* 12.3. DOI: [10.1371/journal.pone.0173372](https://doi.org/10.1371/journal.pone.0173372).
- Lonie, J. et al. (2010). “Predicting outcome in mild cognitive impairment: 4-Year follow-up study”. In: *British Journal of Psychiatry* 197.2, pp. 135–140. DOI: [10.1192/bjp.bp.110.077958](https://doi.org/10.1192/bjp.bp.110.077958).
- López, M. et al. (2016). “Searching for Primary Predictors of Conversion from Mild Cognitive Impairment to Alzheimer’s Disease: A Multivariate Follow-Up Study”. In: *Journal of Alzheimer’s Disease* 52.1, pp. 133–143. DOI: [10.3233/JAD-151034](https://doi.org/10.3233/JAD-151034).
- Mangialasche, F. et al. (2013). “Classification and prediction of clinical diagnosis of Alzheimer’s disease based on MRI and plasma measures of α -/ γ -tocotrienols and γ -tocopherol”. In: *Journal of Internal Medicine* 273.6, pp. 602–621. DOI: [10.1111/joim.12037](https://doi.org/10.1111/joim.12037).
- Marcos, A. et al. (2006). “Neuropsychological markers of progression from mild cognitive impairment to Alzheimer’s disease”. In: *American Journal of Alzheimer’s Disease and other Dementias* 21.3, pp. 189–196. DOI: [10.1177/1533317506289348](https://doi.org/10.1177/1533317506289348).
- Marra, C. et al. (2015). “Usefulness of an Integrated Analysis of Different Memory Tasks to Predict the Progression from Mild Cognitive Impairment to Alzheimer’s Disease: The Episodic Memory Score (EMS)”. In: *Journal of Alzheimer’s Disease* 50.1, pp. 61–70. DOI: [10.3233/JAD-150613](https://doi.org/10.3233/JAD-150613).
- Martinez-Murcia, F. et al. (2018). “Assessing mild cognitive impairment progression using a spherical brain mapping of magnetic resonance imaging”. In: *Journal of Alzheimer’s Disease* 65.3, pp. 713–729. DOI: [10.3233/JAD-170403](https://doi.org/10.3233/JAD-170403).
- Martínez-Torteya, A. et al. (2014). “MRI signal and texture features for the prediction of MCI to Alzheimer’s disease progression”. In: *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*. Vol. 9035, p. 903526. DOI: [10.1117/12.2043903](https://doi.org/10.1117/12.2043903).
- Mathotaarachchi, S. et al. (2017). “Identifying incipient dementia individuals using machine learning and amyloid imaging”. English. In: *Neurobiology of Aging* 0.0. DOI: [10.1016/j.neurobiolaging.2017.06.027](https://doi.org/10.1016/j.neurobiolaging.2017.06.027).
- Mazzeo, S. et al. (2016). “Combining Cerebrospinal Fluid Biomarkers and Neuropsychological Assessment: A Simple and Cost-Effective Algorithm to Predict the Progression from Mild Cognitive Impairment to Alzheimer’s Disease Dementia”. In: *Journal of Alzheimer’s Disease* 54.4, pp. 1495–1508. DOI: [10.3233/JAD-160360](https://doi.org/10.3233/JAD-160360).
- Minhas, S., A. Khanum, F. Riaz, A. Alvi, et al. (2015). “Early alzheimer’s disease prediction in machine learning setup: Empirical analysis with missing value computation”. In: *Lecture Notes in Computer Science (including subseries Lecture*

- Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*) 9375 LNCS, pp. 424–432. DOI: [10.1007/978-3-319-24834-9_49](https://doi.org/10.1007/978-3-319-24834-9_49).
- Minhas, S., A. Khanum, F. Riaz, A. Alvi, et al. (2017). “A Nonparametric Approach for Mild Cognitive Impairment to AD Conversion Prediction: Results on Longitudinal Data”. In: *IEEE Journal of Biomedical and Health Informatics* 21.5, pp. 1403–1410. DOI: [10.1109/JBHI.2016.2608998](https://doi.org/10.1109/JBHI.2016.2608998).
- Minhas, S., A. Khanum, F. Riaz, S. Khan, et al. (2017). “Trajectory based predictive modeling of conversion from mild cognitive impairment to Alzheimer’s disease”. In: *2017 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2017*, pp. 385–388. DOI: [10.1109/BHI.2017.7897286](https://doi.org/10.1109/BHI.2017.7897286).
- (2018). “Predicting progression from mild cognitive impairment to Alzheimer’s disease using autoregressive modelling of longitudinal and multimodal biomarkers”. In: *IEEE Journal of Biomedical and Health Informatics* 22.3, pp. 818–825. DOI: [10.1109/JBHI.2017.2703918](https://doi.org/10.1109/JBHI.2017.2703918).
- Misra, C., Y. Fan, and C. Davatzikos (2009). “Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI”. eng. In: *NeuroImage* 44.4, pp. 1415–1422. DOI: [10.1016/j.neuroimage.2008.10.031](https://doi.org/10.1016/j.neuroimage.2008.10.031).
- Mitchell, J. et al. (2009). “Outcome in subgroups of mild cognitive impairment (MCI) is highly predictable using a simple algorithm”. In: *Journal of Neurology* 256.9, pp. 1500–1509. DOI: [10.1007/s00415-009-5152-0](https://doi.org/10.1007/s00415-009-5152-0).
- Modrego, P. J., N. Fayed, and M. Sarasa (2011). “Magnetic resonance spectroscopy in the prediction of early conversion from amnesic mild cognitive impairment to dementia: a prospective cohort study”. en. In: *BMJ Open* 1.1, e000007–e000007. DOI: [10.1136/bmjopen-2010-000007](https://doi.org/10.1136/bmjopen-2010-000007).
- Modrego, P., N. Fayed, and M. Pina (2005). “Conversion from mild cognitive impairment to probable Alzheimer’s disease predicted by brain magnetic resonance spectroscopy”. In: *American Journal of Psychiatry* 162.4, pp. 667–675. DOI: [10.1176/appi.ajp.162.4.667](https://doi.org/10.1176/appi.ajp.162.4.667).
- Moradi, E. et al. (2015). “Machine learning framework for early MRI-based Alzheimer’s conversion prediction in MCI subjects”. In: *NeuroImage* 104.Supplement C, pp. 398–412. DOI: [10.1016/j.neuroimage.2014.10.002](https://doi.org/10.1016/j.neuroimage.2014.10.002).
- Moretti, D. et al. (2011). “MCI patients’ EEGs show group differences between those who progress and those who do not progress to AD”. In: *Neurobiology of Aging* 32.4, pp. 563–571. DOI: [10.1016/j.neurobiolaging.2009.04.003](https://doi.org/10.1016/j.neurobiolaging.2009.04.003).
- Mubeen, A. et al. (2017). “A six-month longitudinal evaluation significantly improves accuracy of predicting incipient Alzheimer’s disease in mild cognitive impairment”. In: *Journal of Neuroradiology* 44.6, pp. 381–387. DOI: [10.1016/j.neurad.2017.05.008](https://doi.org/10.1016/j.neurad.2017.05.008).

- Nesteruk, M. et al. (2016). “Combined use of biochemical and volumetric biomarkers to assess the risk of conversion of mild cognitive impairment to Alzheimer’s disease”. In: *Folia Neuropathologica* 54.4, pp. 369–374. DOI: [10.5114/fn.2016.64815](https://doi.org/10.5114/fn.2016.64815).
- Nho, K. et al. (2010). “Automatic Prediction of Conversion from Mild Cognitive Impairment to Probable Alzheimer’s Disease using Structural Magnetic Resonance Imaging”. In: *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium 2010*, pp. 542–546.
- Ning, K. et al. (2018). “Classifying Alzheimer’s disease with brain imaging and genetic data using a neural network framework”. In: *Neurobiology of Aging* 68, pp. 151–158. DOI: [10.1016/j.neurobiolaging.2018.04.009](https://doi.org/10.1016/j.neurobiolaging.2018.04.009).
- Ota, K. et al. (2015). “Effects of imaging modalities, brain atlases and feature selection on prediction of Alzheimer’s disease”. In: *Journal of Neuroscience Methods* 256, pp. 168–183. DOI: [10.1016/j.jneumeth.2015.08.020](https://doi.org/10.1016/j.jneumeth.2015.08.020).
- (2016). “Prediction of Alzheimer’s Disease in Amnesic Mild Cognitive Impairment Subtypes: Stratification Based on Imaging Biomarkers”. In: *Journal of Alzheimer’s Disease* 52.4, pp. 1385–1401. DOI: [10.3233/JAD-160145](https://doi.org/10.3233/JAD-160145).
- Papaliagkas, V. et al. (2008). “Usefulness of event-related potentials in the assessment of mild cognitive impairment”. In: *BMC Neuroscience* 9. DOI: [10.1186/1471-2202-9-107](https://doi.org/10.1186/1471-2202-9-107).
- Park, H. et al. (2013). “Dimensionality reduced cortical features and their use in predicting longitudinal changes in Alzheimer’s disease”. In: *Neuroscience Letters* 550, pp. 17–22. DOI: [10.1016/j.neulet.2013.06.042](https://doi.org/10.1016/j.neulet.2013.06.042).
- Parnetti, L. et al. (2012). “Performance of A β 1-40, A β 1-42, total tau, and phosphorylated tau as predictors of dementia in a cohort of patients with mild cognitive impairment”. In: *Journal of Alzheimer’s Disease* 29.1, pp. 229–238. DOI: [10.3233/JAD-2011-111349](https://doi.org/10.3233/JAD-2011-111349).
- Peters, F., S. Villeneuve, and S. Belleville (2014). “Predicting progression to dementia in elderly subjects with mild cognitive impairment using both cognitive and neuroimaging predictors”. In: *Journal of Alzheimer’s Disease* 38.2, pp. 307–318. DOI: [10.3233/JAD-130842](https://doi.org/10.3233/JAD-130842).
- Plant, C. et al. (2010). “Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer’s disease”. In: *NeuroImage* 50.1, pp. 162–174. DOI: [10.1016/j.neuroimage.2009.11.046](https://doi.org/10.1016/j.neuroimage.2009.11.046).
- Plocharski, M. and L. R. Østergaard (2019). “Prediction of Alzheimer’s Disease in Mild Cognitive Impairment Using Sulcal Morphology and Cortical Thickness”. en. In: *World Congress on Medical Physics and Biomedical Engineering 2018. IFMBE Proceedings*. Springer, Singapore, pp. 69–74. ISBN: 978-981-10-9034-9 978-981-10-9035-6. DOI: [10.1007/978-981-10-9035-6_13](https://doi.org/10.1007/978-981-10-9035-6_13).

- Poil, S.-S. et al. (2013). “Integrative EEG biomarkers predict progression to Alzheimer’s disease at the MCI stage”. In: *Frontiers in Aging Neuroscience* 5.OCT. DOI: [10.3389/fnagi.2013.00058](https://doi.org/10.3389/fnagi.2013.00058).
- Popuri, K. et al. (2018). “Development and validation of a novel dementia of Alzheimer’s type (DAT) score based on metabolism FDG-PET imaging”. In: *NeuroImage: Clinical* 18, pp. 802–813. DOI: [10.1016/j.nicl.2018.03.007](https://doi.org/10.1016/j.nicl.2018.03.007).
- Quaranta, D. et al. (2016). “Are raw scores on memory tests better than age- and education- adjusted scores for predicting progression from amnesic mild cognitive impairment to Alzheimer disease?” In: *Current Alzheimer Research* 13.12, pp. 1414–1420. DOI: [10.2174/1567205013666160314145522](https://doi.org/10.2174/1567205013666160314145522).
- Raamana, P. et al. (2015). “Thickness network features for prognostic applications in dementia”. In: *Neurobiology of Aging* 36.S1, S91–S102. DOI: [10.1016/j.neurobiolaging.2014.05.040](https://doi.org/10.1016/j.neurobiolaging.2014.05.040).
- Rabin, L. et al. (2009). “Differential memory test sensitivity for diagnosing amnesic mild cognitive impairment and predicting conversion to Alzheimer’s disease”. In: *Aging, Neuropsychology, and Cognition* 16.3, pp. 357–376. DOI: [10.1080/13825580902825220](https://doi.org/10.1080/13825580902825220).
- Retico, A. et al. (2015). “Predictive Models Based on Support Vector Machines: Whole-Brain versus Regional Analysis of Structural MRI in the Alzheimer’s Disease”. In: *Journal of Neuroimaging* 25.4, pp. 552–563. DOI: [10.1111/jon.12163](https://doi.org/10.1111/jon.12163).
- Rhodijs-Meester, H. et al. (2015). “Integrating Biomarkers for Underlying Alzheimer’s Disease in Mild Cognitive Impairment in Daily Practice: Comparison of a Clinical Decision Support System with Individual Biomarkers”. In: *Journal of Alzheimer’s Disease* 50.1, pp. 261–270. DOI: [10.3233/JAD-150548](https://doi.org/10.3233/JAD-150548).
- Richard, E. et al. (2013). “MRI and cerebrospinal fluid biomarkers for predicting progression to Alzheimer’s disease in patients with mild cognitive impairment: A diagnostic accuracy study”. In: *BMJ Open* 3.6. DOI: [10.1136/bmjopen-2012-002541](https://doi.org/10.1136/bmjopen-2012-002541).
- Ritter, K. et al. (2015). “Multimodal prediction of conversion to Alzheimer’s disease based on incomplete biomarkers”. en. In: *Alzheimer’s & Dementia: Diagnosis, Assessment & Disease Monitoring* 1.2, pp. 206–215. DOI: [10.1016/j.dadm.2015.01.006](https://doi.org/10.1016/j.dadm.2015.01.006).
- Rocha de Paula, M. et al. (2011). “Differences in abundances of cell-signalling proteins in blood reveal novel biomarkers for early detection of clinical Alzheimer’s disease”. In: *PLoS ONE* 6.3. DOI: [10.1371/journal.pone.0017481](https://doi.org/10.1371/journal.pone.0017481).
- Rodriguez, J. et al. (2016). “Short-term MCI-to-AD prediction using MRI, neuropsychological scores and ensemble tree learning techniques”. In: *2015 IEEE Nuclear Science Symposium and Medical Imaging Conference, NSS/MIC 2015*, pp. 1–3. DOI: [10.1109/NSSMIC.2015.7582226](https://doi.org/10.1109/NSSMIC.2015.7582226).

- Sabuncu, M. (2013). “A Bayesian algorithm for image-based time-to-event prediction”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8184 LNCS, pp. 74–81. DOI: [10.1007/978-3-319-02267-3_10](https://doi.org/10.1007/978-3-319-02267-3_10).
- Sanroma, G. et al. (2017). “Early prediction of alzheimer’s disease with non-local patch-based longitudinal descriptors”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10530 LNCS, pp. 74–81. DOI: [10.1007/978-3-319-67434-6_9](https://doi.org/10.1007/978-3-319-67434-6_9).
- Shaffer, J. et al. (2013). “Predicting cognitive decline in subjects at risk for Alzheimer disease by using combined cerebrospinal fluid, MR imaging, and PET biomarkers”. In: *Radiology* 266.2, pp. 583–591. DOI: [10.1148/radiol.12120010](https://doi.org/10.1148/radiol.12120010).
- Shen, T. et al. (2018). “Decision Supporting Model for One-year Conversion Probability from MCI to AD using CNN and SVM”. In: *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*. Vol. 2018, pp. 738–741. DOI: [10.1109/EMBC.2018.8512398](https://doi.org/10.1109/EMBC.2018.8512398).
- Shmulev, Y., M. Belyaev, and A. D. N. I. The (2018). “Predicting conversion of mild cognitive impairments to alzheimer’s disease and exploring impact of neuroimaging”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11044 LNCS, pp. 83–91. DOI: [10.1007/978-3-030-00689-1_9](https://doi.org/10.1007/978-3-030-00689-1_9).
- Si, B., I. Yakushev, and J. Li (2017). “A sequential tree-based classifier for personalized biomarker testing of Alzheimer’s disease risk”. In: *IISE Transactions on Healthcare Systems Engineering* 7.4, pp. 248–260. DOI: [10.1080/24725579.2017.1367979](https://doi.org/10.1080/24725579.2017.1367979).
- Snaedal, J. et al. (2010). “The use of EEG in Alzheimer’s disease, with and without scopolamine - A pilot study”. In: *Clinical Neurophysiology* 121.6, pp. 836–841. DOI: [10.1016/j.clinph.2010.01.008](https://doi.org/10.1016/j.clinph.2010.01.008).
- Sohn, B. et al. (2015). “Comparison of regional gray matter atrophy, white matter alteration, and glucose metabolism as a predictor of the conversion to alzheimer’s disease in mild cognitive impairment”. In: *Journal of Korean Medical Science* 30.6, pp. 779–787. DOI: [10.3346/jkms.2015.30.6.779](https://doi.org/10.3346/jkms.2015.30.6.779).
- Suk, H.-I., S.-W. Lee, and D. Shen (2017). “Deep ensemble learning of sparse regression models for brain disease diagnosis”. In: *Medical Image Analysis* 37, pp. 101–113. DOI: [10.1016/j.media.2017.01.008](https://doi.org/10.1016/j.media.2017.01.008).
- Summers, M. and N. Saunders (2012). “Neuropsychological measures predict decline to alzheimer’s dementia from mild cognitive impairment”. In: *Neuropsychology* 26.4, pp. 498–508. DOI: [10.1037/a0028576](https://doi.org/10.1037/a0028576).

- Tanpitukpongse, T. et al. (2017). “Predictive utility of marketed volumetric software tools in subjects at risk for Alzheimer disease: Do regions outside the hippocampus matter?” In: *American Journal of Neuroradiology* 38.3, pp. 546–552. DOI: [10.3174/ajnr.A5061](https://doi.org/10.3174/ajnr.A5061).
- Teipel, S. et al. (2015). “The relative importance of imaging markers for the prediction of Alzheimer’s disease dementia in mild cognitive impairment - Beyond classical regression”. In: *NeuroImage: Clinical* 8, pp. 583–593. DOI: [10.1016/j.nicl.2015.05.006](https://doi.org/10.1016/j.nicl.2015.05.006).
- Thung, K.-H. et al. (2018). “Conversion and time-to-conversion predictions of mild cognitive impairment using low-rank affinity pursuit denoising and matrix completion”. English. In: *Medical Image Analysis* 45, pp. 68–82. DOI: [10.1016/j.media.2018.01.002](https://doi.org/10.1016/j.media.2018.01.002).
- Tondelli, M. et al. (2015). “Role of cerebrospinal fluid biomarkers to predict conversion to dementia in patients with mild cognitive impairment: A clinical cohort study”. In: *Clinical Chemistry and Laboratory Medicine* 53.3, pp. 453–460. DOI: [10.1515/cclm-2014-0414](https://doi.org/10.1515/cclm-2014-0414).
- Tong, T. and Q. Gao (2015). “Extraction of features from patch based graphs for the prediction of disease progression in AD”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9226, pp. 500–509. DOI: [10.1007/978-3-319-22186-1_50](https://doi.org/10.1007/978-3-319-22186-1_50).
- Tong, T., Q. Gao, et al. (2017). “A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer’s disease”. In: *IEEE Transactions on Biomedical Engineering* 64.1, pp. 155–165. DOI: [10.1109/TBME.2016.2549363](https://doi.org/10.1109/TBME.2016.2549363).
- Trzepacz, P. et al. (2014). “Comparison of neuroimaging modalities for the prediction of conversion from mild cognitive impairment to alzheimer’s dementia”. In: *Neurobiology of Aging* 35.1, pp. 143–151. DOI: [10.1016/j.neurobiolaging.2013.06.018](https://doi.org/10.1016/j.neurobiolaging.2013.06.018).
- Valencia, N. and J. Lehrner (2018). “Screening for dementia with the Vienna Visuo-Constructional Test 3.0 screening (VVT 3.0 screening)”. In: *neuropsychiatrie* 32.4, pp. 196–203.
- Vecchio, D. et al. (2018). “Sustainable method for Alzheimer dementia prediction in mild cognitive impairment: Electroencephalographic connectivity and graph theory combined with apolipoprotein E”. In: *Annals of Neurology* 84.2, pp. 302–314. DOI: [10.1002/ana.25289](https://doi.org/10.1002/ana.25289).
- Visser, P. et al. (1999). “Medial temporal lobe atrophy and memory dysfunction as predictors for dementia in subjects with mild cognitive impairment”. In: *Journal of Neurology* 246.6, pp. 477–485. DOI: [10.1007/s004150050387](https://doi.org/10.1007/s004150050387).

- Vivar, G. et al. (2018). “Multi-modal Disease Classification in Incomplete Datasets Using Geometric Matrix Completion”. In: *arXiv:1803.11550 [cs]*.
- Wang, P. et al. (2016). “Multimodal classification of mild cognitive impairment based on partial least squares”. In: *Journal of Alzheimer’s Disease* 54.1, pp. 359–371. DOI: [10.3233/JAD-160102](https://doi.org/10.3233/JAD-160102).
- Wei, R. et al. (2016). “Prediction of conversion from mild cognitive impairment to Alzheimer’s disease using MRI and structural network features”. In: *Frontiers in Aging Neuroscience* 8.APR. DOI: [10.3389/fnagi.2016.00076](https://doi.org/10.3389/fnagi.2016.00076).
- Westman, E., C. Aguilar, et al. (2013). “Regional magnetic resonance imaging measures for multivariate analysis in Alzheimer’s disease and mild cognitive impairment”. In: *Brain Topography* 26.1, pp. 9–23. DOI: [10.1007/s10548-012-0246-x](https://doi.org/10.1007/s10548-012-0246-x).
- Westman, E., L. Cavallin, et al. (2011). “Sensitivity and specificity of medial temporal lobe visual ratings and multivariate regional MRI classification in Alzheimer’s disease”. In: *PLoS ONE* 6.7. DOI: [10.1371/journal.pone.0022506](https://doi.org/10.1371/journal.pone.0022506).
- Westman, E., J.-S. Muehlboeck, and A. Simmons (2012). “Combining MRI and CSF measures for classification of Alzheimer’s disease and prediction of mild cognitive impairment conversion”. In: *NeuroImage* 62.1, pp. 229–238. DOI: [10.1016/j.neuroimage.2012.04.056](https://doi.org/10.1016/j.neuroimage.2012.04.056).
- Westman, E., A. Simmons, J.-S. Muehlboeck, et al. (2011). “AddNeuroMed and ADNI: Similar patterns of Alzheimer’s atrophy and automated MRI classification accuracy in Europe and North America”. In: *NeuroImage* 58.3, pp. 818–828. DOI: [10.1016/j.neuroimage.2011.06.065](https://doi.org/10.1016/j.neuroimage.2011.06.065).
- Westman, E., A. Simmons, Y. Zhang, et al. (2011). “Multivariate analysis of MRI data for Alzheimer’s disease, mild cognitive impairment and healthy controls”. In: *NeuroImage* 54.2, pp. 1178–1187. DOI: [10.1016/j.neuroimage.2010.08.044](https://doi.org/10.1016/j.neuroimage.2010.08.044).
- Xu, L. et al. (2016). “Prediction of progressive mild cognitive impairment by multimodal neuroimaging biomarkers”. In: *Journal of Alzheimer’s Disease* 51.4, pp. 1045–1056. DOI: [10.3233/JAD-151010](https://doi.org/10.3233/JAD-151010).
- Yang, X., M. Tan, and A. Qiu (2012). “CSF and Brain Structural Imaging Markers of the Alzheimer’s Pathological Cascade”. In: *PLoS ONE* 7.12. DOI: [10.1371/journal.pone.0047406](https://doi.org/10.1371/journal.pone.0047406).
- Ye, D., K. Pohl, and C. Davatzikos (2011). “Semi-supervised pattern classification: Application to structural MRI of Alzheimer’s disease”. In: *Proceedings - International Workshop on Pattern Recognition in NeuroImaging, PRNI 2011*, pp. 1–4. DOI: [10.1109/PRNI.2011.12](https://doi.org/10.1109/PRNI.2011.12).
- Ye, J. et al. (2012). “Sparse learning and stability selection for predicting MCI to AD conversion using baseline ADNI data”. In: *BMC Neurology* 12. DOI: [10.1186/1471-2377-12-46](https://doi.org/10.1186/1471-2377-12-46).

- Young, A. et al. (2014). “A data-driven model of biomarker changes in sporadic Alzheimer’s disease”. In: *Brain* 137.9, pp. 2564–2577. DOI: [10.1093/brain/awu176](https://doi.org/10.1093/brain/awu176).
- Young, J. et al. (2016). “An oblique approach to prediction of conversion to alzheimer’s disease with multikernel gaussian processes”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9444 LNAI, pp. 122–128. DOI: [10.1007/978-3-319-45174-9_13](https://doi.org/10.1007/978-3-319-45174-9_13).
- Yu, P. et al. (2012). “Enriching amnesic mild cognitive impairment populations for clinical trials: Optimal combination of biomarkers to predict conversion to dementia”. In: *Journal of Alzheimer’s Disease* 32.2, pp. 373–385. DOI: [10.3233/JAD-2012-120832](https://doi.org/10.3233/JAD-2012-120832).
- Zhang, D. and D. Shen (2012). “Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer’s disease”. In: *NeuroImage* 59.2, pp. 895–907. DOI: [10.1016/j.neuroimage.2011.09.069](https://doi.org/10.1016/j.neuroimage.2011.09.069).
- Zhang, D., D. Shen, and A. D. N. Initiative (2012). “Predicting Future Clinical Changes of MCI Patients Using Longitudinal and Multimodal Biomarkers”. In: *PLOS ONE* 7.3, e33182. DOI: [10.1371/journal.pone.0033182](https://doi.org/10.1371/journal.pone.0033182).
- Zhang, H. et al. (2018). “Cerebrospinal fluid synaptosomal-associated protein 25 is a key player in synaptic degeneration in mild cognitive impairment and Alzheimer’s disease”. In: *Alzheimer’s Research and Therapy* 10.1. DOI: [10.1186/s13195-018-0407-6](https://doi.org/10.1186/s13195-018-0407-6).
- Zhao, Y. et al. (2017). “Predicting MCI progression with individual metabolic network based on longitudinal FDG-PET”. In: *Proceedings - 2017 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2017*. Vol. 2017-January, pp. 1894–1899. DOI: [10.1109/BIBM.2017.8217948](https://doi.org/10.1109/BIBM.2017.8217948).
- Zheng, W. et al. (2018). “Identification of Alzheimer’s Disease and Mild Cognitive Impairment Using Networks Constructed Based on Multiple Morphological Brain Features”. en. In: *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 3.10, pp. 887–897. DOI: [10.1016/j.bpsc.2018.06.004](https://doi.org/10.1016/j.bpsc.2018.06.004).
- Zhu, X., H.-I. Suk, and D. Shen (2014a). “A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis”. In: *NeuroImage* 100, pp. 91–105. DOI: [10.1016/j.neuroimage.2014.05.078](https://doi.org/10.1016/j.neuroimage.2014.05.078).
- (2014b). “A novel multi-relation regularization method for regression and classification in AD diagnosis”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8675 LNCS.PART 3, pp. 401–408. DOI: [10.1007/978-3-319-10443-0_51](https://doi.org/10.1007/978-3-319-10443-0_51).

Zhu, X., H.-I. Suk, L. Wang, et al. (2017). “A novel relational regularization feature selection method for joint regression and classification in AD diagnosis”. In: *Medical Image Analysis* 38, pp. 205–214. DOI: [10.1016/j.media.2015.10.008](https://doi.org/10.1016/j.media.2015.10.008).

Bibliography

- AKSMAN, L. M. (2017). « Longitudinal neuroimaging features for discriminating early neurodegeneration ». en. Thèse de doct. King's College London.
- AMIEVA, H., H. JACQMIN-GADDA et al. (2005). « The 9 year cognitive decline before dementia of the Alzheimer type : a prospective population-based study ». en. In : *Brain* 128.5, p. 1093–1101. DOI : [10.1093/brain/awh451](https://doi.org/10.1093/brain/awh451).
- AMIEVA, H., H. MOKRI et al. (2014). « Compensatory mechanisms in higher-educated subjects with Alzheimer's disease : a study of 20 years of cognitive decline ». en. In : *Brain* 137.4, p. 1167–1175. DOI : [10.1093/brain/awu035](https://doi.org/10.1093/brain/awu035).
- ANSART, M., S. EPELBAUM, G. GAGLIARDI, O. COLLIOT, D. DORMONT, B. DU-BOIS et al. (2019). « Reduction of recruitment costs in preclinical AD trials : validation of automatic pre-screening algorithm for brain amyloidosis ». en. In : *Statistical Methods in Medical Research*, p. 096228021882303. DOI : [10.1177/0962280218823036](https://doi.org/10.1177/0962280218823036).
- ANSART, M., S. EPELBAUM, G. GAGLIARDI, O. COLLIOT, D. DORMONT, H. HAMPPEL et al. (2017). « Prediction of Amyloidosis from Neuropsychological and MRI Data for Cost Effective Inclusion of Pre-symptomatic Subjects in Clinical Trials ». en. In : *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. T. 10553. Cham : Springer International Publishing, p. 357–364. ISBN : 978-3-319-67557-2 978-3-319-67558-9. DOI : [10.1007/978-3-319-67558-9_41](https://doi.org/10.1007/978-3-319-67558-9_41).
- ANTILA, K. et al. (2013). « The PredictAD project : development of novel biomarkers and analysis software for early diagnosis of the Alzheimer's disease ». In : *Interface Focus* 3.2. DOI : [10.1098/rsfs.2012.0072](https://doi.org/10.1098/rsfs.2012.0072).
- APOSTOLOVA, L. G. et al. (2015). « Brain amyloidosis ascertainment from cognitive, imaging, and peripheral blood protein measures ». In : *Neurology* 84.7, p. 729–737. DOI : [10.1212/WNL.0000000000001231](https://doi.org/10.1212/WNL.0000000000001231).
- ARBABSHIRANI, M. R. et al. (2017). « Single subject prediction of brain disorders in neuroimaging : Promises and pitfalls ». In : *NeuroImage*. Individual Subject Prediction 145, p. 137–165. DOI : [10.1016/j.neuroimage.2016.02.079](https://doi.org/10.1016/j.neuroimage.2016.02.079).
- ARDEKANI, B. A. et al. (2017). « Prediction of Incipient Alzheimer's Disease Dementia in Patients with Mild Cognitive Impairment ». en. In : *Journal of Alzheimer's Disease* 55.1, p. 269–281. DOI : [10.3233/JAD-160594](https://doi.org/10.3233/JAD-160594).

- ATRAMONT, A. et al. (2018). « Health status and drug use 1 year before and 1 year after skilled nursing home admission during the first quarter of 2013 in France : a study based on the French National Health Insurance Information System ». eng. In : *European Journal of Clinical Pharmacology* 74.1, p. 109–118. DOI : [10.1007/s00228-017-2343-y](https://doi.org/10.1007/s00228-017-2343-y).
- BARNES, D. E. et K. YAFFE (2011). « The Projected Impact of Risk Factor Reduction on Alzheimer's Disease Prevalence ». In : *Lancet neurology* 10.9, p. 819–828. DOI : [10.1016/S1474-4422\(11\)70072-2](https://doi.org/10.1016/S1474-4422(11)70072-2).
- BATEMAN, R. J. et al. (2012). « Clinical and Biomarker Changes in Dominantly Inherited Alzheimer's Disease ». en. In : *New England Journal of Medicine* 367.9, p. 795–804. DOI : [10.1056/NEJMoa1202753](https://doi.org/10.1056/NEJMoa1202753).
- BEACH, T. G. et al. (2012). « Accuracy of the Clinical Diagnosis of Alzheimer Disease at National Institute on Aging Alzheimer's Disease Centers, 2005–2010 ». In : *Journal of Neuropathology and Experimental Neurology* 71.4, p. 266–273. DOI : [10.1097/NEN.0b013e31824b211b](https://doi.org/10.1097/NEN.0b013e31824b211b).
- BECKER, R. E. et N. H. GREIG (2014). « A New Regulatory Road-Map for Alzheimer's Disease Drug Development ». In : *Current Alzheimer research* 11.3, p. 215–220.
- BEHESHTI, I., H. DEMIREL et H. MATSUDA (2017). « Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm ». en. In : *Computers in Biology and Medicine* 83, p. 109–119. DOI : [10.1016/j.combiomed.2017.02.011](https://doi.org/10.1016/j.combiomed.2017.02.011).
- BIÉTRY, F. A. et al. (2017). « Benzodiazepine Use and Risk of Developing Alzheimer's Disease : A Case-Control Study Based on Swiss Claims Data ». en. In : *CNS Drugs* 31.3, p. 245–251. DOI : [10.1007/s40263-016-0404-x](https://doi.org/10.1007/s40263-016-0404-x).
- BILLIOTI DE GAGE, S. et al. (2014). « Benzodiazepine use and risk of Alzheimer's disease : case-control study ». eng. In : *BMJ (Clinical research ed.)* 349, g5205. DOI : [10.1136/bmj.g5205](https://doi.org/10.1136/bmj.g5205).
- BIRKS, J. S. et J. GRIMLEY EVANS (2015). « Rivastigmine for Alzheimer's disease ». eng. In : *The Cochrane Database of Systematic Reviews* 4, p. CD001191. DOI : [10.1002/14651858.CD001191.pub3](https://doi.org/10.1002/14651858.CD001191.pub3).
- BREIMAN, L. (2001). « Random forests ». In : *Machine learning* 45.1, p. 5–32.
- CHENG, G. et al. (2012). « Diabetes as a risk factor for dementia and mild cognitive impairment : a meta-analysis of longitudinal studies ». eng. In : *Internal Medicine Journal* 42.5, p. 484–491. DOI : [10.1111/j.1445-5994.2012.02758.x](https://doi.org/10.1111/j.1445-5994.2012.02758.x).
- CHÉTELAT, G. et al. (2013). « Amyloid imaging in cognitively normal individuals, at-risk populations and preclinical Alzheimer's disease ». In : *NeuroImage : Clinical* 2, p. 356–365. DOI : [10.1016/j.nicl.2013.02.006](https://doi.org/10.1016/j.nicl.2013.02.006).

- CHIARA, G. D. et al. (2019). « Recurrent herpes simplex virus-1 infection induces hallmarks of neurodegeneration and cognitive deficits in mice ». en. In : *PLOS Pathogens* 15.3, e1007617. DOI : [10.1371/journal.ppat.1007617](https://doi.org/10.1371/journal.ppat.1007617).
- CHINCARINI, A. et al. (2011). « Local MRI analysis approach in the diagnosis of early and prodromal Alzheimer's disease ». In : *NeuroImage* 58.2, p. 469–480. DOI : [10.1016/j.neuroimage.2011.05.083](https://doi.org/10.1016/j.neuroimage.2011.05.083).
- CHOI, H. et K. JIN (2018). « Predicting cognitive decline with deep learning of brain metabolism and amyloid imaging ». In : *Behavioural Brain Research* 344, p. 103–109. DOI : [10.1016/j.bbr.2018.02.017](https://doi.org/10.1016/j.bbr.2018.02.017).
- CHUPIN, M. et al. (2009). « Automatic segmentation of the hippocampus and the amygdala driven by hybrid constraints : method and validation ». eng. In : *NeuroImage* 46.3, p. 749–761. DOI : [10.1016/j.neuroimage.2009.02.013](https://doi.org/10.1016/j.neuroimage.2009.02.013).
- CLARK, C. M. et al. (2012). « Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid- β plaques : a prospective cohort study ». In : *The Lancet Neurology* 11.8, p. 669–678. DOI : [10.1016/S1474-4422\(12\)70142-4](https://doi.org/10.1016/S1474-4422(12)70142-4).
- CUI, Y. et al. (2011). « Identification of Conversion from Mild Cognitive Impairment to Alzheimer's Disease Using Multivariate Predictors ». In : *PLOS ONE* 6.7, e21896. DOI : [10.1371/journal.pone.0021896](https://doi.org/10.1371/journal.pone.0021896).
- CUINGNET, R. et al. (2011). « Automatic classification of patients with Alzheimer's disease from structural MRI : A comparison of ten methods using the ADNI database ». en. In : *NeuroImage* 56.2, p. 766–781. DOI : [10.1016/j.neuroimage.2010.06.013](https://doi.org/10.1016/j.neuroimage.2010.06.013).
- DAVATZIKOS, C. et al. (2011). « Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification ». In : *Neurobiology of Aging* 32.12, 2322.e19–2322.e27. DOI : [10.1016/j.neurobiolaging.2010.05.023](https://doi.org/10.1016/j.neurobiolaging.2010.05.023).
- DING, Y. et al. (2018). « A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using 18F-FDG PET of the Brain ». In : *Radiology* 290.2, p. 456–464. DOI : [10.1148/radiol.2018180958](https://doi.org/10.1148/radiol.2018180958).
- DOMINGOS, P. (2012). « A few useful things to know about machine learning ». en. In : *Communications of the ACM* 55.10, p. 78. DOI : [10.1145/2347736.2347755](https://doi.org/10.1145/2347736.2347755).
- DONEGAN, K. et al. (2017). « Trends in diagnosis and treatment for people with dementia in the UK from 2005 to 2015 : a longitudinal retrospective cohort study ». eng. In : *The Lancet. Public Health* 2.3, e149–e156. DOI : [10.1016/S2468-2667\(17\)30031-2](https://doi.org/10.1016/S2468-2667(17)30031-2).
- DOODY, R. S. et al. (2014). « Phase 3 Trials of Solanezumab for Mild-to-Moderate Alzheimer's Disease ». In : *New England Journal of Medicine* 370.4, p. 311–321. DOI : [10.1056/NEJMoa1312889](https://doi.org/10.1056/NEJMoa1312889).

- DOYLE, O. et al. (2014). « Predicting progression of Alzheimer's disease using ordinal regression ». In : *PLoS ONE* 9.8. DOI : [10.1371/journal.pone.0105542](https://doi.org/10.1371/journal.pone.0105542).
- DUBOIS, B., H. H. FELDMAN, C. JACOVA, S. T. DEKOSKY et al. (2007). « Research criteria for the diagnosis of Alzheimer's disease : revising the NINCDS-ADRDA criteria ». eng. In : *The Lancet. Neurology* 6.8, p. 734–746. DOI : [10.1016/S1474-4422\(07\)70178-3](https://doi.org/10.1016/S1474-4422(07)70178-3).
- DUBOIS, B., H. H. FELDMAN, C. JACOVA, H. HAMPEL et al. (2014). « Advancing research diagnostic criteria for Alzheimer's disease : the IWG-2 criteria ». English. In : *The Lancet Neurology* 13.6, p. 614–629. DOI : [10.1016/S1474-4422\(14\)70090-0](https://doi.org/10.1016/S1474-4422(14)70090-0).
- DUBOIS, B., H. HAMPEL et al. (2016). « Preclinical Alzheimer's disease : Definition, natural history, and diagnostic criteria ». In : *Alzheimer's & Dementia* 12.3, p. 292–323. DOI : [10.1016/j.jalz.2016.02.002](https://doi.org/10.1016/j.jalz.2016.02.002).
- DUKART, J., F. SAMBATARO et A. BERTOLINO (2015). « Accurate prediction of conversion to Alzheimer's disease using imaging, genetic, and neuropsychological biomarkers ». In : *Journal of Alzheimer's Disease* 49.4, p. 1143–1159. DOI : [10.3233/JAD-150570](https://doi.org/10.3233/JAD-150570).
- EFRON, B. (1983). « Estimating the Error Rate of a Prediction Rule : Improvement on Cross-Validation ». In : *Journal of the American Statistical Association* 78.382, p. 316–331. DOI : [10.1080/01621459.1983.10477973](https://doi.org/10.1080/01621459.1983.10477973).
- EPELBAUM, S., R. GENTHON et al. (2017). « Preclinical Alzheimer's disease : A systematic review of the cohorts underlying the concept ». English. In : *Alzheimer's & Dementia : The Journal of the Alzheimer's Association* 13.4, p. 454–467. DOI : [10.1016/j.jalz.2016.12.003](https://doi.org/10.1016/j.jalz.2016.12.003).
- EPELBAUM, S., C. PAQUET et al. (2019). « How many patients are eligible for disease-modifying treatment in Alzheimer's disease ? A French national observational study over 5 years ». en. In : *BMJ Open* 9.6, e029663. DOI : [10.1136/bmjopen-2019-029663](https://doi.org/10.1136/bmjopen-2019-029663).
- ESCOTT-PRICE, V. et al. (2015). « Common polygenic variation enhances risk prediction for Alzheimer's disease ». In : *Brain* 138.12, p. 3673–3684. DOI : [10.1093/brain/awv268](https://doi.org/10.1093/brain/awv268).
- ESKILDSEN, S. et al. (2015). « Structural imaging biomarkers of Alzheimer's disease : Predicting disease progression ». In : *Neurobiology of Aging* 36.S1, S23–S31. DOI : [10.1016/j.neurobiolaging.2014.04.034](https://doi.org/10.1016/j.neurobiolaging.2014.04.034).
- FALAHATI, F., E. WESTMAN et A. SIMMONS (2014). « Multivariate Data Analysis and Machine Learning in Alzheimer's Disease with a Focus on Structural Magnetic Resonance Imaging ». en. In : *Journal of Alzheimer's Disease* 41.3, p. 685–708. DOI : [10.3233/JAD-131928](https://doi.org/10.3233/JAD-131928).

- FLEISHER, A. et al. (2008). « Volumetric MRI vs clinical predictors of Alzheimer disease in mild cognitive impairment ». In : *Neurology* 70.3, p. 191–199. DOI : [10.1212/01.wnl.0000287091.57376.65](https://doi.org/10.1212/01.wnl.0000287091.57376.65).
- FOSTER, J. C., J. M. TAYLOR et S. J. RUBERG (2011). « Subgroup identification from randomized clinical trial data ». en. In : *Statistics in Medicine* 30.24, p. 2867–2880. DOI : [10.1002/sim.4322](https://doi.org/10.1002/sim.4322).
- FRANKE, K. et al. (2010). « Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods : Exploring the influence of various parameters ». In : *NeuroImage* 50.3, p. 883–892. DOI : [10.1016/j.neuroimage.2010.01.005](https://doi.org/10.1016/j.neuroimage.2010.01.005).
- FRIEDMAN, J., T. HASTIE et R. TIBSHIRANI (2000). « Additive logistic regression : a statistical view of boosting ». In : *The annals of statistics* 28.2, p. 337–407.
- FRIEDMAN, J. H. (2001). « Greedy Function Approximation : A Gradient Boosting Machine ». In : *The Annals of Statistics* 29.5, p. 1189–1232.
- FRIEDMAN, J., T. HASTIE et R. TIBSHIRANI (2010). « Regularization paths for generalized linear models via coordinate descent ». In : *Journal of statistical software* 33.1, p. 1.
- GAVIDIA-BOVADILLA, G. et al. (2017). « Early prediction of Alzheimer’s disease using null longitudinal model-based classifiers ». In : *PLoS ONE* 12.1. DOI : [10.1371/journal.pone.0168011](https://doi.org/10.1371/journal.pone.0168011).
- GÓMEZ-SANCHO, M., J. TOHKA et V. GÓMEZ-VERDEJO (2018). « Comparison of feature representations in MRI-based MCI-to-AD conversion prediction ». en. In : *Magnetic Resonance Imaging* 50, p. 84–95. DOI : [10.1016/j.mri.2018.03.003](https://doi.org/10.1016/j.mri.2018.03.003).
- GRAY, K. R. et al. (2012). « Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer’s disease ». In : *NeuroImage* 60.1, p. 221–229. DOI : [10.1016/j.neuroimage.2011.12.071](https://doi.org/10.1016/j.neuroimage.2011.12.071).
- GRAY, S. L. et al. (2016). « Benzodiazepine use and risk of incident dementia or cognitive decline : prospective population based study ». eng. In : *BMJ (Clinical research ed.)* 352, p. i90. DOI : [10.1136/bmj.i90](https://doi.org/10.1136/bmj.i90).
- HALL, A. et al. (2015). « Predicting progression from cognitive impairment to Alzheimer’s disease with the disease state index ». In : *Current Alzheimer Research* 12.1, p. 69–79. DOI : [10.2174/1567205012666141218123829](https://doi.org/10.2174/1567205012666141218123829).
- HALLER, S., K. O. LOVBLAD et P. GIANNAKOPOULOS (2011). « Principles of Classification Analyses in Mild Cognitive Impairment (MCI) and Alzheimer Disease ». en. In : *Journal of Alzheimer’s Disease* 26.s3. Sous la dir. de J. W. ASHFORD et al., p. 389–394. DOI : [10.3233/JAD-2011-0014](https://doi.org/10.3233/JAD-2011-0014).
- HAMELIN, L. et al. (2016). « Early and protective microglial activation in Alzheimer’s disease : a prospective study using 18F-DPA-714 PET imaging ». eng. In :

- Brain : A Journal of Neurology* 139.Pt 4, p. 1252–1264. DOI : [10.1093/brain/aww017](https://doi.org/10.1093/brain/aww017).
- HAND, D. J. et R. J. TILL (2001). « A simple generalisation of the area under the ROC curve for multiple class classification problems ». In : *Machine learning* 45.2, p. 171–186.
- HARDY, J. A. et G. A. HIGGINS (1992). « Alzheimer's disease : the amyloid cascade hypothesis ». In : *Science* 256.5054, p. 184–185.
- HARDY, J. et D. ALLSOP (1991). « Amyloid deposition as the central event in the aetiology of Alzheimer's disease ». In : *Trends in Pharmacological Sciences* 12, p. 383–388. DOI : [10.1016/0165-6147\(91\)90609-V](https://doi.org/10.1016/0165-6147(91)90609-V).
- HARDY, J. et D. J. SELKOE (2002). « The Amyloid Hypothesis of Alzheimer's Disease : Progress and Problems on the Road to Therapeutics ». en. In : *Science* 297.5580, p. 353–356. DOI : [10.1126/science.1072994](https://doi.org/10.1126/science.1072994).
- HARIRFOROOSH, S., W. ASGHAR et F. JAMALI (2013). « Adverse effects of nonsteroidal antiinflammatory drugs : an update of gastrointestinal, cardiovascular and renal complications ». eng. In : *Journal of Pharmacy & Pharmaceutical Sciences : A Publication of the Canadian Society for Pharmaceutical Sciences, Societe Canadienne Des Sciences Pharmaceutiques* 16.5, p. 821–847.
- HARRIS, S. A. et E. A. HARRIS (2018). « Molecular Mechanisms for Herpes Simplex Virus Type 1 Pathogenesis in Alzheimer's Disease ». eng. In : *Frontiers in Aging Neuroscience* 10, p. 48. DOI : [10.3389/fnagi.2018.00048](https://doi.org/10.3389/fnagi.2018.00048).
- HOERL, A. E. et R. W. KENNARD (1970). « Ridge Regression : Biased Estimation for Nonorthogonal Problems ». In : *Technometrics* 12.1, p. 55–67. DOI : [10.1080/00401706.1970.10488634](https://doi.org/10.1080/00401706.1970.10488634).
- HOLMES, C. (2013). « Review : systemic inflammation and Alzheimer's disease ». eng. In : *Neuropathology and Applied Neurobiology* 39.1, p. 51–68. DOI : [10.1111/j.1365-2990.2012.01307.x](https://doi.org/10.1111/j.1365-2990.2012.01307.x).
- HUGHES, G. (1968). « On the mean accuracy of statistical pattern recognizers ». In : *IEEE Transactions on Information Theory* 14.1, p. 55–63. DOI : [10.1109/TIT.1968.1054102](https://doi.org/10.1109/TIT.1968.1054102).
- HYVARINEN, A. (1999). « Fast and robust fixed-point algorithms for independent component analysis ». In : *IEEE transactions on Neural Networks* 10.3, p. 626–634.
- IDDI, S. et al. (2019). « Predicting the course of Alzheimer's progression ». en. In : *Brain Informatics* 6.1, p. 6. DOI : [10.1186/s40708-019-0099-0](https://doi.org/10.1186/s40708-019-0099-0).
- INSEL, P. S. et al. (2016). « Assessing risk for preclinical β -amyloid pathology with APOE, cognitive, and demographic information ». In : *Alzheimer's & Dementia : Diagnosis, Assessment & Disease Monitoring* 4, p. 76–84.

- JACK, C. R., D. A. BENNETT et al. (2018). « NIA-AA Research Framework : Toward a biological definition of Alzheimer's disease ». en. In : *Alzheimer's & Dementia* 14.4, p. 535–562. DOI : [10.1016/j.jalz.2018.02.018](https://doi.org/10.1016/j.jalz.2018.02.018).
- JACK, C. R., D. S. KNOPMAN et al. (2010). « Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade ». In : *Lancet neurology* 9.1, p. 119. DOI : [10.1016/S1474-4422\(09\)70299-6](https://doi.org/10.1016/S1474-4422(09)70299-6).
- JACK, C. R., H. J. WISTE et al. (2015). « Different definitions of neurodegeneration produce similar amyloid/neurodegeneration biomarker group findings ». en. In : *Brain* 138.12, p. 3747–3759. DOI : [10.1093/brain/awv283](https://doi.org/10.1093/brain/awv283).
- JATURAPATPORN, D. et al. (2012). « Aspirin, steroidal and non-steroidal anti-inflammatory drugs for the treatment of Alzheimer's disease ». eng. In : *The Cochrane Database of Systematic Reviews* 2, p. CD006378. DOI : [10.1002/14651858.CD006378.pub2](https://doi.org/10.1002/14651858.CD006378.pub2).
- JOHNSON, P. et al. (2014). « Genetic algorithm with logistic regression for prediction of progression to Alzheimer's disease ». In : *BMC Bioinformatics* 15. DOI : [10.1186/1471-2105-15-S16-S11](https://doi.org/10.1186/1471-2105-15-S16-S11).
- JUNWEI DING et QIU HUANG (2017). « Prediction of MCI to AD conversion using Laplace Eigenmaps learned from FDG and MRI images of AD patients and healthy controls ». en. In : *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*. Chengdu, China : IEEE, p. 660–664. ISBN : 978-1-5090-6238-6. DOI : [10.1109/ICIVC.2017.7984638](https://doi.org/10.1109/ICIVC.2017.7984638).
- KARRAN, E., M. MERCKEN et B. D. STROOPER (2011). « The amyloid cascade hypothesis for Alzheimer's disease : an appraisal for the development of therapeutics ». en. In : *Nature Reviews Drug Discovery* 10.9, p. 698–712. DOI : [10.1038/nrd3505](https://doi.org/10.1038/nrd3505).
- KATE, M. ten et al. (2018). « MRI predictors of amyloid pathology : results from the EMIF-AD Multimodal Biomarker Discovery study ». In : *Alzheimer's Research & Therapy* 10.1, p. 100. DOI : [10.1186/s13195-018-0428-1](https://doi.org/10.1186/s13195-018-0428-1).
- KAUPPI, K. et al. (2018). « Combining polygenic hazard score with volumetric MRI and cognitive measures improves prediction of progression from mild cognitive impairment to Alzheimer's disease ». In : *Frontiers in Neuroscience* 12.APR. DOI : [10.3389/fnins.2018.00260](https://doi.org/10.3389/fnins.2018.00260).
- KIM, J. et B. LEE (2018). « Identification of Alzheimer's disease and mild cognitive impairment using multimodal sparse hierarchical extreme learning machine ». en. In : *Human Brain Mapping*. DOI : [10.1002/hbm.24207](https://doi.org/10.1002/hbm.24207).
- KING, E. et al. (2018). « Peripheral inflammation in prodromal Alzheimer's and Lewy body dementias ». eng. In : *Journal of Neurology, Neurosurgery, and Psychiatry* 89.4, p. 339–345. DOI : [10.1136/jnnp-2017-317134](https://doi.org/10.1136/jnnp-2017-317134).

- KISHI, T. et al. (2017). « Memantine for Alzheimer's Disease : An Updated Systematic Review and Meta-analysis ». eng. In : *Journal of Alzheimer's disease : JAD* 60.2, p. 401–425. DOI : [10.3233/JAD-170424](https://doi.org/10.3233/JAD-170424).
- KOROLEV, I. O. et al. (2016). « Predicting Progression from Mild Cognitive Impairment to Alzheimer's Dementia Using Clinical, MRI, and Plasma Biomarkers via Probabilistic Pattern Classification ». en. In : *PLOS ONE* 11.2, e0138866. DOI : [10.1371/journal.pone.0138866](https://doi.org/10.1371/journal.pone.0138866).
- KOVAL, I., S. ALLASSONNIÈRE et S. DURRLEMAN (2019). « Simulation of virtual cohorts increases predictive accuracy of cognitive decline in MCI subjects ». In : *arXiv :1904.02921 [cs, stat]*.
- KOVAL, I., A. BÔNE et al. (2018). « Simulating Alzheimer's disease progression with personalised digital brain models ».
- LAUZON, F. Q. (2012). « An introduction to deep learning ». en. In : *2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA)*. Montreal, QC, Canada : IEEE, p. 1438–1439. ISBN : 978-1-4673-0382-8 978-1-4673-0381-1 978-1-4673-0380-4. DOI : [10.1109/ISSPA.2012.6310529](https://doi.org/10.1109/ISSPA.2012.6310529).
- LEANDROU, S. et al. (2018). « Quantitative MRI Brain Studies in Mild Cognitive Impairment and Alzheimer's Disease : A Methodological Review ». In : *IEEE Reviews in Biomedical Engineering* 11, p. 97–111. DOI : [10.1109/RBME.2018.2796598](https://doi.org/10.1109/RBME.2018.2796598).
- LEI, B. et al. (2016). « Discriminative learning for Alzheimer's disease diagnosis via canonical correlation analysis and multimodal fusion ». In : *Frontiers in Aging Neuroscience* 8.MAY. DOI : [10.3389/fnagi.2016.00077](https://doi.org/10.3389/fnagi.2016.00077).
- LI DENG (2012). « The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web] ». en. In : *IEEE Signal Processing Magazine* 29.6, p. 141–142. DOI : [10.1109/MSP.2012.2211477](https://doi.org/10.1109/MSP.2012.2211477).
- LI, Y. et al. (2018). « Indirect relation based individual metabolic network for identification of mild cognitive impairment ». In : *Journal of Neuroscience Methods* 309, p. 188–198. DOI : [10.1016/j.jneumeth.2018.09.007](https://doi.org/10.1016/j.jneumeth.2018.09.007).
- LIN, W.-Y. et al. (2019). « Association of Antiviral Therapy With Risk of Parkinson Disease in Patients With Chronic Hepatitis C Virus Infection ». eng. In : *JAMA neurology*. DOI : [10.1001/jamaneurol.2019.1368](https://doi.org/10.1001/jamaneurol.2019.1368).
- LIN, W. et al. (2018). « Convolutional neural networks-based MRI image analysis for the Alzheimer's disease prediction from mild cognitive impairment ». In : *Frontiers in Neuroscience* 12.NOV. DOI : [10.3389/fnins.2018.00777](https://doi.org/10.3389/fnins.2018.00777).
- LOPEZ, C. L. et al. (2017). « The Alzheimer's Prevention Initiative Generation Program : Evaluating CNP520 Efficacy in the Prevention of Alzheimer's Disease. » In : *The journal of prevention of Alzheimer's disease* 4.4, p. 242–246.

- MARENGONI, A. et al. (2009). « Patterns of Chronic Multimorbidity in the Elderly Population ». en. In : *Journal of the American Geriatrics Society* 57.2, p. 225–230. DOI : [10.1111/j.1532-5415.2008.02109.x](https://doi.org/10.1111/j.1532-5415.2008.02109.x).
- MARINESCU, R. V. et al. (2018). « TADPOLE Challenge : Prediction of Longitudinal Evolution in Alzheimer's Disease ». In : *arXiv :1805.03909 [q-bio, stat]*.
- MARTIN, C. et al. (2014). « Inflammatory and neurodegeneration markers during asymptomatic HSV-1 reactivation ». eng. In : *Journal of Alzheimer's disease : JAD* 39.4, p. 849–859. DOI : [10.3233/JAD-131706](https://doi.org/10.3233/JAD-131706).
- MASSOT MESQUIDA, M. et al. (2019). « Consensus and evidence-based medication review to optimize and potentially reduce psychotropic drug prescription in institutionalized dementia patients ». In : *BMC Geriatrics* 19.1, p. 7. DOI : [10.1186/s12877-018-1015-9](https://doi.org/10.1186/s12877-018-1015-9).
- MENDES, A. et al. (2018). « Multimorbidity Is Associated with Preclinical Alzheimer's Disease Neuroimaging Biomarkers ». eng. In : *Dementia and Geriatric Cognitive Disorders* 45.5-6, p. 272–281. DOI : [10.1159/000489007](https://doi.org/10.1159/000489007).
- MIELKE, M. M. et al. (2012). « Indicators of amyloid burden in a population-based study of cognitively normal elderly ». In : *Neurology* 79.15, p. 1570–1577. DOI : [10.1212/WNL.0b013e31826e2696](https://doi.org/10.1212/WNL.0b013e31826e2696).
- MINHAS, S. et al. (2018). « Predicting progression from mild cognitive impairment to Alzheimer's disease using autoregressive modelling of longitudinal and multimodal biomarkers ». In : *IEEE Journal of Biomedical and Health Informatics* 22.3, p. 818–825. DOI : [10.1109/JBHI.2017.2703918](https://doi.org/10.1109/JBHI.2017.2703918).
- MINSKER, S., Y.-Q. ZHAO et G. CHENG (2016). « Active Clinical Trials for Personalized Medicine ». en. In : *Journal of the American Statistical Association* 111.514, p. 875–887. DOI : [10.1080/01621459.2015.1066682](https://doi.org/10.1080/01621459.2015.1066682).
- MORADI, E. et al. (2015). « Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects ». In : *NeuroImage* 104.Supplement C, p. 398–412. DOI : [10.1016/j.neuroimage.2014.10.002](https://doi.org/10.1016/j.neuroimage.2014.10.002).
- MUBEEN, A. et al. (2017). « A six-month longitudinal evaluation significantly improves accuracy of predicting incipient Alzheimer's disease in mild cognitive impairment ». In : *Journal of Neuroradiology* 44.6, p. 381–387. DOI : [10.1016/j.neurad.2017.05.008](https://doi.org/10.1016/j.neurad.2017.05.008).
- MULLER, K. R. et al. (2001). « An introduction to kernel-based learning algorithms ». In : *IEEE Transactions on Neural Networks* 12.2, p. 181–201. DOI : [10.1109/72.914517](https://doi.org/10.1109/72.914517).
- NAKAMURA, A. et al. (2018). « High performance plasma amyloid- β biomarkers for Alzheimer's disease ». en. In : *Nature* 554.7691, p. 249–254. DOI : [10.1038/nature25456](https://doi.org/10.1038/nature25456).

- NHO, K. et al. (2010). « Automatic Prediction of Conversion from Mild Cognitive Impairment to Probable Alzheimer's Disease using Structural Magnetic Resonance Imaging ». In : *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium 2010*, p. 542–546.
- NORTON, S. et al. (2014). « Potential for primary prevention of Alzheimer's disease : an analysis of population-based data ». eng. In : *The Lancet. Neurology* 13.8, p. 788–794. DOI : [10.1016/S1474-4422\(14\)70136-X](https://doi.org/10.1016/S1474-4422(14)70136-X).
- O'BRIEN, J. T. et K. HERHOLZ (2015). « Amyloid imaging for dementia in clinical practice ». In : *BMC Medicine* 13, p. 163–169. DOI : [10.1186/s12916-015-0404-6](https://doi.org/10.1186/s12916-015-0404-6).
- ORGANIZATION, W. H. (2017). *Herpes simplex virus*. en.
- OTA, K. et al. (2015). « Effects of imaging modalities, brain atlases and feature selection on prediction of Alzheimer's disease ». In : *Journal of Neuroscience Methods* 256, p. 168–183. DOI : [10.1016/j.jneumeth.2015.08.020](https://doi.org/10.1016/j.jneumeth.2015.08.020).
- PERERA, G. et al. (2014). « Factors associated with response to acetylcholinesterase inhibition in dementia : a cohort study from a secondary mental health care case register in london ». eng. In : *PloS One* 9.11, e109484. DOI : [10.1371/journal.pone.0109484](https://doi.org/10.1371/journal.pone.0109484).
- PETERSEN, R. C. et al. (2013). « Mild cognitive impairment due to Alzheimer disease in the community ». eng. In : *Annals of Neurology* 74.2, p. 199–208. DOI : [10.1002/ana.23931](https://doi.org/10.1002/ana.23931).
- PRINCE, M., D. R. BRYCE et D. C. FERRI (2011). « World Alzheimer Report 2011 : The benefits of early diagnosis and intervention ». en. In : *Alzheimer's Disease International*, p. 72.
- PRINCE, M., A. WIMO et al. (2015). « World Alzheimer Report 2015, The Global Impact of Dementia : An analysis of prevalence, incidence, cost and trends ». en. In : *Alzheimer's Disease International*, p. 87.
- QIAN, M. et S. A. MURPHY (2011). « Performance guarantees for individualized treatment rules ». en. In : *The Annals of Statistics* 39.2, p. 1180–1210. DOI : [10.1214/10-AOS864](https://doi.org/10.1214/10-AOS864).
- RAAMANA, P. et al. (2015). « Thickness network features for prognostic applications in dementia ». In : *Neurobiology of Aging* 36.S1, S91–S102. DOI : [10.1016/j.neurobiolaging.2014.05.040](https://doi.org/10.1016/j.neurobiolaging.2014.05.040).
- RATHORE, S. et al. (2017). « A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages ». en. In : *NeuroImage* 155, p. 530–548. DOI : [10.1016/j.neuroimage.2017.03.057](https://doi.org/10.1016/j.neuroimage.2017.03.057).
- RENOM-GUITERAS, A. et al. (2018). « Potentially inappropriate medication among people with dementia in eight European countries ». eng. In : *Age and Ageing* 47.1, p. 68–74. DOI : [10.1093/ageing/afx147](https://doi.org/10.1093/ageing/afx147).

- RETICO, A. et al. (2015). « Predictive Models Based on Support Vector Machines : Whole-Brain versus Regional Analysis of Structural MRI in the Alzheimer's Disease ». In : *Journal of Neuroimaging* 25.4, p. 552–563. DOI : [10.1111/jon.12163](https://doi.org/10.1111/jon.12163).
- RICHARDSON, K. et al. (2018). « Anticholinergic drugs and risk of dementia : case-control study ». en. In : *BMJ* 361, k1315. DOI : [10.1136/bmj.k1315](https://doi.org/10.1136/bmj.k1315).
- SABUNCU, M. (2013). « A Bayesian algorithm for image-based time-to-event prediction ». In : *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8184 LNCS, p. 74–81. DOI : [10.1007/978-3-319-02267-3_10](https://doi.org/10.1007/978-3-319-02267-3_10).
- SALLOWAY, S. et al. (2014). « Two Phase 3 Trials of Bapineuzumab in Mild-to-Moderate Alzheimer's Disease ». In : *New England Journal of Medicine* 370.4, p. 322–333. DOI : [10.1056/NEJMoa1304839](https://doi.org/10.1056/NEJMoa1304839).
- SAMPER-GONZALEZ, J. et al. (2019). « Reproducible evaluation of methods for predicting progression to Alzheimer's disease from clinical and neuroimaging data ». en. In : *Medical Imaging 2019 : Image Processing*. Sous la dir. d'E. D. ANGELINI et B. A. LANDMAN. San Diego, United States : SPIE, p. 30. ISBN : 978-1-5106-2545-7 978-1-5106-2546-4. DOI : [10.1117/12.2512430](https://doi.org/10.1117/12.2512430).
- SAMPER-GONZÁLEZ, J. et al. (2018). « Reproducible evaluation of classification methods in Alzheimer's disease : framework and application to MRI and PET data ». en. In : *NeuroImage* 183, p. 504–521. DOI : [10.1101/274324](https://doi.org/10.1101/274324).
- SARICA, A., A. CERASA et A. QUATTRONE (2017). « Random Forest Algorithm for the Classification of Neuroimaging Data in Alzheimer's Disease : A Systematic Review ». In : *Frontiers in Aging Neuroscience* 9, p. 329. DOI : [10.3389/fnagi.2017.00329](https://doi.org/10.3389/fnagi.2017.00329).
- SATLIN, A. et al. (2016). « Design of a Bayesian adaptive phase 2 proof-of-concept trial for BAN2401, a putative disease-modifying monoclonal antibody for the treatment of Alzheimer's disease ». en. In : *Alzheimer's & Dementia : Translational Research & Clinical Interventions* 2.1, p. 1–12. DOI : [10.1016/j.trci.2016.01.001](https://doi.org/10.1016/j.trci.2016.01.001).
- SCHUSTER, C. et al. (2015). « Presymptomatic and longitudinal neuroimaging in neurodegeneration—from snapshots to motion picture : a systematic review ». en. In : *Journal of Neurology, Neurosurgery & Psychiatry* 86.10, p. 1089–1096. DOI : [10.1136/jnnp-2014-309888](https://doi.org/10.1136/jnnp-2014-309888).
- SHAFFER, J. et al. (2013). « Predicting cognitive decline in subjects at risk for Alzheimer disease by using combined cerebrospinal fluid, MR imaging, and PET biomarkers ». In : *Radiology* 266.2, p. 583–591. DOI : [10.1148/radiol.12120010](https://doi.org/10.1148/radiol.12120010).

- SHAW, L. M. et al. (2009). « Cerebrospinal Fluid Biomarker Signature in Alzheimer's Disease Neuroimaging Initiative Subjects ». In : *Annals of neurology* 65.4, p. 403–413. DOI : [10.1002/ana.21610](https://doi.org/10.1002/ana.21610).
- SI, B., I. YAKUSHEV et J. LI (2017). « A sequential tree-based classifier for personalized biomarker testing of Alzheimer's disease risk ». In : *IISE Transactions on Healthcare Systems Engineering* 7.4, p. 248–260. DOI : [10.1080/24725579.2017.1367979](https://doi.org/10.1080/24725579.2017.1367979).
- SPERLING, R. A. et al. (2011). « Toward defining the preclinical stages of Alzheimer's disease : Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease ». In : *Alzheimer's & dementia : the journal of the Alzheimer's Association* 7.3, p. 280–292. DOI : [10.1016/j.jalz.2011.03.003](https://doi.org/10.1016/j.jalz.2011.03.003).
- TANPITUKPONGSE, T. et al. (2017). « Predictive utility of marketed volumetric software tools in subjects at risk for Alzheimer disease : Do regions outside the hippocampus matter ? ». In : *American Journal of Neuroradiology* 38.3, p. 546–552. DOI : [10.3174/ajnr.A5061](https://doi.org/10.3174/ajnr.A5061).
- THUNG, K.-H. et al. (2018). « Conversion and time-to-conversion predictions of mild cognitive impairment using low-rank affinity pursuit denoising and matrix completion ». English. In : *Medical Image Analysis* 45, p. 68–82. DOI : [10.1016/j.media.2018.01.002](https://doi.org/10.1016/j.media.2018.01.002).
- TONG, T. et al. (2017). « A novel grading biomarker for the prediction of conversion from mild cognitive impairment to Alzheimer's disease ». In : *IEEE Transactions on Biomedical Engineering* 64.1, p. 155–165. DOI : [10.1109/TBME.2016.2549363](https://doi.org/10.1109/TBME.2016.2549363).
- TOSUN, D., S. JOSHI et M. W. WEINER (2013). « Neuroimaging Predictors of Brain Amyloidosis in Mild Cognitive Impairment ». In : *Annals of neurology* 74.2, p. 188–198. DOI : [10.1002/ana.23921](https://doi.org/10.1002/ana.23921).
- TZENG, N.-S. et al. (2018). « Anti-herpetic Medications and Reduced Risk of Dementia in Patients with Herpes Simplex Virus Infections-a Nationwide, Population-Based Cohort Study in Taiwan ». eng. In : *Neurotherapeutics : The Journal of the American Society for Experimental NeuroTherapeutics* 15.2, p. 417–429. DOI : [10.1007/s13311-018-0611-x](https://doi.org/10.1007/s13311-018-0611-x).
- VECCHIO, D. et al. (2018). « Sustainable method for Alzheimer dementia prediction in mild cognitive impairment : Electroencephalographic connectivity and graph theory combined with apolipoprotein E ». In : *Annals of Neurology* 84.2, p. 302–314. DOI : [10.1002/ana.25289](https://doi.org/10.1002/ana.25289).
- VERMUNT, L. et al. (2019). « Duration of preclinical, prodromal, and dementia stages of Alzheimer's disease in relation to age, sex, and APOE genotype ». English. In : *Alzheimer's & Dementia : The Journal of the Alzheimer's Association* 15.7, p. 888–898. DOI : [10.1016/j.jalz.2019.04.001](https://doi.org/10.1016/j.jalz.2019.04.001).

- VIVAR, G. et al. (2018). « Multi-modal Disease Classification in Incomplete Datasets Using Geometric Matrix Completion ». In : *arXiv :1803.11550 [cs]*.
- WATSON, J. L. et al. (2014). « Obstacles And Opportunities In Alzheimer's Clinical Trial Recruitment ». In : *Health affairs (Project Hope)* 33.4, p. 574–579. DOI : [10.1377/hlthaff.2013.1314](https://doi.org/10.1377/hlthaff.2013.1314).
- WEN, J. et al. (2019). « How serious is data leakage in deep learning studies on Alzheimer's disease classification? » en. In : *Proceedings of : Human Brain Mapping (HBM)*, p. 8.
- WESTMAN, E., C. AGUILAR et al. (2013). « Regional magnetic resonance imaging measures for multivariate analysis in Alzheimer's disease and mild cognitive impairment ». In : *Brain Topography* 26.1, p. 9–23. DOI : [10.1007/s10548-012-0246-x](https://doi.org/10.1007/s10548-012-0246-x).
- WESTMAN, E., J.-S. MUEHLBOECK et A. SIMMONS (2012). « Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion ». In : *NeuroImage* 62.1, p. 229–238. DOI : [10.1016/j.neuroimage.2012.04.056](https://doi.org/10.1016/j.neuroimage.2012.04.056).
- What is Alzheimer's?* (2019).
- WHITMER, R. A. et al. (2005). « Midlife cardiovascular risk factors and risk of dementia in late life ». eng. In : *Neurology* 64.2, p. 277–281. DOI : [10.1212/01.WNL.0000149519.47454.F2](https://doi.org/10.1212/01.WNL.0000149519.47454.F2).
- WITTE, M. M. et al. (2015). « Clinical use of amyloid-positron emission tomography neuroimaging : Practical and bioethical considerations ». In : *Alzheimer's & Dementia : Diagnosis, Assessment & Disease Monitoring* 1.3, p. 358–367. DOI : [10.1016/j.dadm.2015.06.006](https://doi.org/10.1016/j.dadm.2015.06.006).
- WOLPERT, D. H. (2002). « The Supervised Learning No-Free-Lunch Theorems ». en. In : *Soft Computing and Industry*. Sous la dir. de R. ROY et al. London : Springer London, p. 25–42. ISBN : 978-1-4471-1101-6 978-1-4471-0123-9. DOI : [10.1007/978-1-4471-0123-9_3](https://doi.org/10.1007/978-1-4471-0123-9_3).
- WOLPERT, D. H. et W. G. MACREADY (1997). « No Free Lunch Theorems for Optimization ». en. In : *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION* 1.1, p. 16.
- WOOLLER, S. (2018). *Artificial intelligence can predict Alzheimer's 6 years earlier than medics, study finds*. en-US. Text.Article.
- WOZNIAK, M. A. et al. (2007). « Herpes simplex virus infection causes cellular beta-amyloid accumulation and secretase upregulation ». eng. In : *Neuroscience Letters* 429.2-3, p. 95–100. DOI : [10.1016/j.neulet.2007.09.077](https://doi.org/10.1016/j.neulet.2007.09.077).
- YE, D., K. POHL et C. DAVATZIKOS (2011). « Semi-supervised pattern classification : Application to structural MRI of Alzheimer's disease ». In : *Proceedings -*

International Workshop on Pattern Recognition in NeuroImaging, PRNI 2011, p. 1–4. DOI : [10.1109/PRNI.2011.12](https://doi.org/10.1109/PRNI.2011.12).

ZHAO, Y. et al. (2012). « Estimating Individualized Treatment Rules Using Outcome Weighted Learning ». en. In : *Journal of the American Statistical Association* 107.499, p. 1106–1118. DOI : [10.1080/01621459.2012.695674](https://doi.org/10.1080/01621459.2012.695674).