Tampere University

HILKKA LIEDES

Prediction and Monitoring of Progression of Alzheimer's Disease

Multivariable approaches for decision support

Tampere University Dissertations 736

Tampere University Dissertations 736

HILKKA LIEDES

Prediction and Monitoring of Progression of Alzheimer's Disease

Multivariable approaches for decision support

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Medicine and Health Technology
of Tampere University,
for public discussion in the auditorium TB109
of the Tietotalo building, Korkeakoulunkatu 7, Tampere,
on 17 February 2023, at 12 o'clock.

ACADEMIC DISSERTATION
Tampere University, Faculty of Medicine and Health Technology
VTT Technical Research Centre of Finland
Finland

Responsible supervisor and Custos	Professor Mark van Gils Tampere University Finland	
Supervisor	Adjunct Professor Ilkka Korhonen Tampere University Finland	
Pre-examiners	Professor Tapio Seppänen University of Oulu Finland	Professor Mikael von und zu Fraunberg University of Oulu Finland
Opponent	Professor Natasha Maurits University of Groningen Netherlands	

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Cover design: Roihu Inc.

ISBN 978-952-03-2734-7 (print)
ISBN 978-952-03-2735-4 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
http://urn.fi/URN:ISBN:978-952-03-2735-4

Carbon dioxide emissions from printing Tampere University dissertations have been compensated.

PunaMusta Oy – Yliopistopaino
Joensuu 2023

PREFACE

The research presented in this thesis was conducted at VTT Technical Research Centre of Finland Ltd during the years 2012-2022. This thesis would not have seen the light of day without the support from many great people who I have had the privilege to work with. I want to express my gratitude to several key persons.

First and foremost, I am extremely grateful to my supervisor Professor Mark van Gils. You have been guiding my researcher career since the hiring process and my first day at VTT. I highly appreciate your expertise in this field and your willingness to share it; it has been a great pleasure to work with you during these years. Your warm-heartedness and kindness have made it easy to come to you with all sorts of questions and problems – thank you for having your door always open.

I owe my thanks to Adjunct Professor Jyrki Lötjönen who guided the practical research work on all four publications forming this thesis. Jyrki, I admire your dedication to this research topic; and I want to thank you for obtaining funding for and coordinating Alzheimer's disease related research projects at VTT. Your rigorous and precise review of analysis results, all questions, and constructive comments on the article drafts were crucial for this work.

I want to express my thanks to the members of my thesis follow-up group, Adjunct Professor Ilkka Korhonen and Dr. Harri Pölönen. I am grateful to Ilkka for being also my supervisor and guiding my PhD journey and thesis writing. My warm thanks to Harri for reviewing the thesis, sharing an office room with me, and having countless lunch and coffee breaks together.

I am very grateful to all co-authors of the articles. Professor Wiesje van der Flier and Dr. Hanneke Rhodius-Meester, it was a great pleasure to collaborate with you in the PredictND project and your clinical expertise is very much appreciated. Wiesje, thank you for supervising two studies leading to two articles of this thesis. Hanneke, many, many thanks for your work for the two studies – this thesis project would have been so much more difficult without it. Professor Emerita Hilkka Soininen, I want to thank you for your contributions for the "longitudinal Disease State Index" article. Dr. Mark Gordon and Dr. Gerald Novak, thank you for your contributions to the "hippocampal atrophy" article. I want to express my gratitude to my (ex-)VTT colleagues, Adjunct Professor Juha Koikkalainen, Dr. Jussi Mattila, and

Juha Kortelainen for being awesome colleagues and for their technical work for the articles. Finally, I want to acknowledge all the data owners, who provided us with an access to their data. Possibility to use data sets from different countries made the articles a lot stronger.

I am very grateful to my pre-examiners Professor Tapio Seppänen and Professor Mikael von und zu Fraunberg for the thorough evaluation of my thesis as well as their kind and constructive feedback. I also wish to thank Professor Natasha Maurits for agreeing to act as my opponent in the public defence of this thesis.

I want to thank my current and former colleagues at VTT for the discussions, laughs, humour, and relaxing lunch and coffee breaks. It has always been easy to ask all kinds of questions and there has always been people willing to help. Thank you for the great atmosphere at the workplace.

This work would not have been possible without funding received. I acknowledge the funding from the European Union Seventh Framework Programme, the Finnish Funding Agency for Technology and Innovation (TEKES), and the Innovative Medicines Initiative Joint Undertaking (IMI-JU), resources of which are composed of financial contribution from the European Union's Seventh Framework Programme and the European Federation of Pharmaceutical Industries and Associations' (EFPIA) in kind contribution. The thesis was finalized with funding from VTT Technical Research Centre of Finland Ltd and Tampere University.

On a personal level, I am grateful to my family and friends. My parents, Eeva and Paavo, thank you for providing me with great foundations for my life. I am grateful for all your love and support since the childhood. My siblings Heikki, Esa, and Soile, thank you for all common activities and trips with you and your families during these years. Pedro, big thanks to you for your peer support during my PhD journey, dance activities, and life in general. It has been a privilege to get to know you and your family. Heidi, Anna, Tiina, and Kata, thank you for your company and support during our master's studies and afterwards. Sari, thank you for your friendship that has lasted since our childhood - it means a lot to me. Finally, I want to express my deepest and warmest gratitude to my little family. My beloved Petri, thank you for your love, support, and patience with this whole thing. Sharing my life and all the experiences with you has been priceless. Our little Aapo, you have brought so much joy and love in my life. You both are my everything.

Oulu, November 2022
Hilkka Liedes

ABSTRACT

Alzheimer's disease (AD), the most common form of dementia, is a slowly progressing neurodegenerative disease, which cannot be cured yet. However, certain medications and lifestyle interventions can delay progression of the disease and its symptoms, thereby positively influencing both quality of life of patients as well as cost-effectiveness of healthcare. Early diagnosis of AD is important because such interventions should be started already at an early phase of the disease to have the best effect. However, early diagnosis is challenging because pathological changes in the brain occur years before the clinical symptoms become visible. In addition, the research during the past years has produced information from a large number of different tests and biomarkers that can potentially contribute to diagnosis and prognosis of AD. This excessive amount of data can cause information overload for clinicians, thus hampering the clinicians' decision making. Data-driven analysis and visualization methods may help with interpretation and utilization of large amounts of heterogeneous patient data and support the clinicians' decision-making process. Furthermore, the methods may aid in identifying suitable patients for clinical drug trials.

The aim of the work described in this thesis was to develop and validate data-driven methods for predicting and monitoring progression of Alzheimer's disease at the different phases of the disease spectrum, starting from normal cognition and ending to death, using data from neuropsychological and cognitive tests, magnetic resonance imaging (MRI), cerebrospinal fluid samples (CSF), comorbidities, and genetics (apolipoprotein E).

The thesis consists of four original studies published as international journal articles. The first study focused on the early phase of AD. A supervised machine learning method called Disease State Index (DSI) was utilized to predict who of the individuals with subjective cognitive decline (SCD) will progress to a more severe condition, i.e., mild cognitive impairment (MCI) or dementia. The study population included 647 subjects from three different memory clinic-based cohorts in Europe. When all data modalities were combined, the area under the receiver operating characteristic curve (AUC) was 0.81 and balanced accuracy was 74%. Negative predictive value was high (93%), whereas positive predictive value was low (38%). Performance

of the DSI method in terms of AUC decreased by 11% when validated with an independent test set. Additional analyses suggested that several differences between the cohorts may explain the decrease in the performance.

The second study focused on a more advanced disease stage. The DSI method was applied to longitudinal data collected from an MCI cohort of 273 subjects obtained from the Alzheimer's Disease and Neuroimaging (ADNI 1) study. Longitudinal profiles of the DSI values differed between the subjects progressing to dementia due to AD and subjects remaining as MCI. In addition, two subgroups were found in the group remaining as MCI: one group with stable DSI values over time and another group with increasing DSI values, suggesting the latter group may progress to dementia due to AD in the future. This study also extended the Disease State Fingerprint (DSF) data visualization method for longitudinal data.

The third study predicted hippocampal atrophy over 24 months using baseline data and penalized linear regression. The cohorts consisted of subjects with normal cognition, MCI, and dementia due to AD and were obtained from the ADNI 1 (n=530) and the Australian Imaging Biomarkers and Lifestyle Flagship Study of Ageing (AIBL, n=176) studies. The models including different data modalities performed better than the models including only MRI features. However, both models underestimated the real change at higher atrophy rate levels, the MRI-only models showing a greater underestimation. When predicting dichotomized outcome, i.e., fast vs. slow atrophy, the models obtained a prediction accuracy of 79-87%. The MRI-only models performed well when evaluated with an independent validation cohort (AIBL).

The last study focused on the latest phase of AD by identifying which disease-related determinants are associated with mortality in patients with dementia due to AD. The cohort included 616 patients from the Amsterdam Dementia Cohort. Age- and sex-adjusted Cox proportional hazards models revealed that older age, male sex, and worse scores on cognitive functioning, as well as more severe medial temporal lobe and global cortical atrophy were associated with an increased risk of mortality. An optimal combination of variables comprised age, sex, performance on digit span backward test and Trail Making Test A, medial temporal lobe atrophy, and tau phosphorylated at threonine 181 in CSF.

In conclusion, data-driven methods can be used for predicting and monitoring progression of AD from the mildest stages to the more advanced stages. Combining information from several data modalities provides better prediction performance than individual data modalities alone. The results also highlight the importance of the validation of the methods with independent validation cohorts. Introduction of

these methods to different environments and countries may require harmonization of patient examination methods and diagnostic criteria.

Keywords: Alzheimer's disease, clinical decision support, data visualization, dementia, machine learning, multivariable modelling, prediction, supervised learning

TIIVISTELMÄ

Alzheimerin tauti, yksi yleisimmistä muistisairauksista, on hitaasti etenevä aivoja rappeuttava tauti, jolle ei ole vielä parantavaa hoitoa. Tietyt lääkkeet ja elämäntapainterventiot voivat kuitenkin hidastaa taudin etenemistä ja lievittää sen oireita, mikä parantaa potilaiden elämänlaatua ja terveydenhuollon kustannusvaikuttavuutta. Alzheimerin taudin varhainen diagnostiikka on erittäin tärkeää, koska erilaiset interventiot pitäisi aloittaa jo taudin varhaisessa vaiheessa, jotta niillä saataisiin aikaan paras mahdollinen vaikutus. Taudin varhainen diagnostiikka on kuitenkin haastavaa, koska muutokset aivoissa alkavat vuosia tai vuosikymmeniä ennen ensimmäisten oireiden ilmaantumista. Lisäksi viime vuosien tutkimus on tuottanut tietoa suuresta määrästä erilaisia testejä ja biomarkkereita, jotka voivat vaikuttaa taudin diagnoosiin ja prognoosiin. Tiedon suuri määrä saattaa aiheuttaa informaatioähkyä kliinikoille vaikeuttaen heidän päätöksentekoaan. Datalähtöiset analytiikka- ja visualisointimenetelmät voivat auttaa suuren ja heterogeenisen tietomäärän tulkinnassa ja hyödyntämisessä. Ne voivat siten tukea kliinikkoa hänen päätöksenteossaan. Lisäksi nämä menetelmät voivat auttaa tunnistamaan sopivia potilaita kliinisiin lääketutkimuksiin, joiden tavoitteena on kehittää Alzheimerin taudin etenemistä hidastavia lääkkeitä.

Tämän väitöskirjan tavoitteena oli kehittää datalähtöisiä menetelmiä Alzheimerin taudin etenemisen ennustamiseen ja seurantaan taudin eri vaiheisiin alkaen normaalista kognitiosta ja edeten kuolemaan. Mallien kehittämisessä hyödynnettiin kognitiivisten ja neuropsykologisten testien tuloksia, magneettikuvantamista (MRI), selkäydinnestenäytteitä, ja genetiikkaa (apolipoproteiini E).

Väitöskirja koostuu neljästä alkuperäisestä tutkimuksesta, jotka on julkaistu kansainvälisissä tieteellisissä lehdissä. Ensimmäinen osatutkimus keskittyi Alzheimerin taudin varhaiseen vaiheeseen. Tutkimuksessa käytettiin ohjattua koneoppimisen menetelmää Disease State Index (DSI, taudin tilan indeksi) ennustamaan, kenellä subjektiivisesti koettu kognition heikkeneminen etenee taudin vakavampaan vaiheeseen eli lievään kognition heikentymiseen (mild cognitive impairment, MCI) tai dementiaan. Tutkimuksen aineisto koostui 647 henkilöstä kolmesta eurooppalaisesta muistiklinikkakohortista. Kun yhdistettiin useita eri muuttujia DSI-menetelmällä, ROC-käyrän (engl. Receiver Operating Characteristic curve) alle jäävä pinta-ala (AUC) oli 0.81 ja tasapainotettu tarkkuus oli 74%. Negatiivinen ennustearvo oli korkea (93%)

ja positiivinen ennustearvo oli matala (38%). Kun DSI-malli validoitiin erillisellä testikohortilla, mallin AUC huononi 11%. Lisäanalyysit osoittivat, että useat erot kohorttien välillä voivat selittää suorituskyvyn alenemista.

Toinen osatutkimus keskittyi taudin myöhäisempään vaiheeseen. DSI-menetelmällä analysoitiin pitkittäistä dataa, joka koostui 273 henkilön MCI-kohortista. Kohortti hankittiin Alzheimer's Disease and Neuroimaging (ADNI 1) tietokannasta. DSI-arvojen muutokset ajan kuluessa olivat erilaiset niillä, joiden tauti eteni Alzheimerin taudin dementiaksi, ja niillä, joilla tauti pysyi MCI-vaiheessa. Lisäksi huomattiin, että stabiilina pysynyt MCI-ryhmä koostui kahdesta aliryhmästä: ensimmäisessä ryhmässä DSI-arvot pysyivät vakaina ja toisessa ryhmässä DSI-arvot kohosivat. Tämä indikoi, että toisessa ryhmässä tauti saattaa edetä dementiaksi tulevaisuudessa. Näiden analyysien lisäksi DSI:in oleellisesti liittyvä Disease State Fingerprint (DSF, taudin tilan sormenjälki) -visualisointimenetelmä laajennettiin pitkittäiselle datalle.

Kolmas osatutkimus ennusti hippokampuksen surkastumista 24 kuukauden aikana lähtötilanteen mittausten perusteella. Tutkimuskohortti koostui henkilöistä, joilla oli normaali kognitio, MCI tai Alzheimerin taudin dementia, ja se hankittiin ADNI 1 (n=530) ja Australian Imaging Biomarkers and Lifestyle Flagship Study of Ageing (AIBL, n=176) tutkimuksista. Useita eri datatyyppejä sisältävät mallit ennustivat hippokampuksen surkastumista tarkemmin kuin pelkistä MRI-muuttujista koostuvat mallit. Kuitenkin molemmat mallit aliarvioivat todellista surkastumista erityisesti suuremmilla surkastumisnopeuksilla, aliarviointi oli suurempaa pelkästään MRI-muuttujiin perustuvilla malleilla. Kun ennustettiin kaksiluokkaista vastemuuttujaa, eli nopea vs. hidas surkastuminen, mallien tarkkuus oli 79-87%. MRI-mallien suorituskyky oli hyvä, kun testauksessa käytettiin erillistä AIBL-aineistoa.

Viimeinen osatutkimus keskittyi Alzheimerin taudin viimeisimpiin vaiheisiin. Siinä tutkittiin, mitkä tautiin liittyvät tekijät ovat yhteydessä kuolleisuuteen potilailla, joilla oli Alzheimerin taudin dementia. Aineisto koostui 616 henkilöstä Amsterdam Dementia Cohort -aineistosta. Iällä ja sukupuolella vakioidun Coxin suhteellisen vaaran mallin mukaan vanhempi ikä, miessukupuoli, huonommat pisteet kognitiivisessa toimintakyvyssä, ja aivojen kuoriosien ja mediaalisen ohimolohkon surkastuminen olivat yhteydessä kuolleisuuteen. Optimaalinen muuttujien yhdistelmä sisälsi iän, sukupuolen, tulokset kahdesta kognitiivisesta testistä (digit span backward, Trail Making Test A), mediaalisen ohimolohkon surkastumisen ja selkäydinnestenäytteestä mitatun kohdasta 181 (treoniini) fosforyloidun tau-proteiinin määrän.

Yhteenvetona todetaan, että datalähtöisillä menetelmillä voidaan ennustaa ja seurata Alzheimerin taudin etenemistä varhaisesta vaiheesta myöhäiseen vaiheeseen. Yhdistämällä useita eri datatyyppejä saadaan parempia tuloksia kuin käyttämällä vain

yhtä datatyyppiä. Tulokset korostavat myös, että datalähtöiset menetelmät on tärkeä arvioida erillisellä aineistolla, jota ei ole käytetty menetelmien kehittämiseen. Lisäksi näiden menetelmien käyttöönotto eri ympäristöissä tai maissa saattaa vaatia potilaan tutkimusmenetelmien ja diagnoosikriteereiden harmonisointia.

Avainsanat: Alzheimerin tauti, datan visualisointi, dementia, ennustaminen, kliininen päätöksenteontuki, koneoppiminen, monimuuttujamallinnus, ohjattu oppiminen

ORIGINAL COMMUNICATIONS

This thesis is based on the following original publications which are referred to in the text as I–IV. The publications are ordered here by the severity of the Alzheimer's disease starting from an early phase and ending to a dementia phase and death. Name of the author of this thesis (Liedes, previously Runtti) is in bold in the publication list. Publication I is reproduced under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International license (https://creativecommons.org/licenses/by-nc-nd/4.0/). Publications II and III are reproduced with the kind permission from the publisher. Publication IV is reproduced under Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0). The figures obtained from the original publications are reproduced in this thesis under these agreements.

I Rhodius-Meester, H. F. M, **Liedes, H.**, Koikkalainen, J., Wolfsgruber, S., Coll-Padros, N., Kornhuber, J., Peters, O., Jessen, F., Kleineidam, L., Molinuevo, J. L., Rami, L., Teunissen, C. E., Barkhof, F., Sikkes, S. A. M., Wesselman, L. M. P., Slot, R. E. R., Verfaillie, S. C. J., Scheltens, P., Tijms, B. M., Lötjönen, J. & van der Flier, W. M. (2018) Computer assisted prediction of clinical progression in the earliest stages of AD. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 10, 726-736. https://doi.org/10.1016/j.dadm.2018.09.001.

II **Runtti, H.**, Mattila, J., van Gils, M., Koikkalainen, J., Soininen, H., Lötjönen, J. & for the Alzheimer's Disease Neuroimaging Initiative (2014) Quantitative evaluation of disease progression in a longitudinal mild cognitive impairment cohort. *Journal of Alzheimer's Disease*, 39(1), 49-61. https://doi.org/10.3233/JAD-130359.

III **Liedes, H.**, Lötjönen, J., Kortelainen, J. M., Novak, G., van Gils, M., Gordon, M. F. & for the Alzheimer's Disease Neuroimaging Initiative & the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (2019) Multivariate prediction of hippocampal atrophy in Alzheimer's disease. *Journal of Alzheimer's Disease*, 68(4), 1453-1468. https://doi.org/10.3233/JAD-180484.

IV Rhodius-Meester, H. F. M., **Liedes, H.**, Koene, T., Lemstra, A. W., Teunissen, C. E., Barkhof, F., Scheltens, P., van Gils, M., Lötjönen, J. & van der Flier, W. M. (2018) Disease-related determinants are associated with mortality in dementia due to Alzheimer's disease. *Alzheimer's Research & Therapy*, 10(23). https://doi.org/10.1186/s13195-018-0348-0.

AUTHOR'S CONTRIBUTION

Author's contributions to the four original publications were as follows:

I The author conducted machine learning analysis, interpreted the results, and participated in the manuscript writing. H. F. M. Rhodius-Meester was the main author of the publication and analysed the baseline characteristics. J. Koikkalainen extracted the MRI features.

II The author analysed the data, interpreted the results, and had the main responsibility for the writing the publication. J. Mattila and J. Lötjönen designed the method and J. Mattila implemented it. The study was designed by J. Lötjönen, J. Mattila, and the author of the thesis.

III The author analysed the data, interpreted the results, and was the main author of the publication. J. Lötjönen calculated the annual hippocampal atrophy rates and J. Kortelainen extracted the MRI features at the baseline.

IV The author conducted preliminary data analysis, interpreted the data, and participated in the manuscript writing. H. F. M. Rhodius-Meester had the main responsibility for the data analysis and publication writing.

CONTENTS

Preface iii

Abstract v

Tiivistelmä viii

Original communications xi

Author's contribution xii

Contents xiii

Symbols and abbreviations xvii

1 Introduction 21
- 1.1 Background 21
- 1.2 Outline of the thesis 22

2 Review of the literature 24
- 2.1 Dementia 24
- 2.2 Alzheimer's disease 25
- 2.3 The continuum of Alzheimer's disease 26
  - 2.3.1 Pathological changes of the brain in Alzheimer's disease 26
  - 2.3.2 Biomarkers of Alzheimer's disease 27
  - 2.3.3 Risk factors of Alzheimer's disease 28
  - 2.3.4 Temporal evolution of the biomarkers in Alzheimer's disease 29
  - 2.3.5 Diagnostic guidelines for Alzheimer's disease 31
  - 2.3.6 Risk factors for mortality in Alzheimer's disease 33
- 2.4 Clinical decision support systems for dementia 34
- 2.5 Development of machine learning systems 36
  - 2.5.1 Design cycle of machine learning systems 36
  - 2.5.2 Performance evaluation of machine learning systems 38
  - 2.5.3 Performance metrics used in the evaluation 39
- 2.6 Machine learning in Alzheimer's disease 41
  - 2.6.1 Typical research questions 41

2.6.2 Data sources and data modalities....42
2.6.3 Developed models....43
2.6.4 Performance of the models....44
2.6.5 Challenges and limitations....49

3 Aims of the study....51

4 Materials and methods....52
4.1 Study populations and variables....52
4.1.1 Prediction of progression of SCD (Study I)....53
4.1.2 Monitoring progression of MCI (Study II)....54
4.1.3 Prediction of hippocampal atrophy (Study III)....55
4.1.4 Factors associated with mortality in AD (Study IV)....56
4.2 Modelling methods....57
4.2.1 Disease State Index....57
4.2.2 Disease State Fingerprint....59
4.2.3 Naïve Bayes....60
4.2.4 Random Forests....62
4.2.5 Regularized regression....62
4.2.6 Least absolute deviation regression....63
4.2.7 Cox proportional hazards model....63
4.3 Performance evaluation....64

5 Results....67
5.1 Prediction of progression of SCD (Study I)....67
5.2 Monitoring progression of MCI (Study II)....72
5.3 Prediction of hippocampal atrophy (Study III)....75
5.4 Factors associated with mortality in AD (Study IV)....79

6 Discussion....82
6.1 Accomplishment of the objectives....82
6.2 Comparison to prior work....84
6.3 Impact of the research in its field....88
6.4 Limitations of the studies....91
6.5 Future work....92

7 Conclusions....96

References....98

List of Figures

Figure 1. Panels A and B: A hypothetical model describing temporal evolution of the major Alzheimer's disease biomarkers and cognitive impairment which is shown as a light green area bounded by limits for high and low risk. Panel B: Operational use of the model. The black vertical line denotes a given time (T) and the horizontal dashed arrows show values of each biomarker at time T. The grey circles highlight that patients with similar biomarker profiles at time T can have different levels of cognitive impairment, depending on their risk factors for AD (Jack et al., 2013).

Figure 2. A general design cycle of a machine learning system (based on Duda et al. (2001) and Webb and Copsey (2011))

Figure 3. Confusion matrix

Figure 4. Summary of the objectives

Figure 5. An example of a Disease State Fingerprint visualization.

Figure 6. Examples of the Disease State Fingerprint visualizations: patients A and C remained stable and patient B progressed to MCI.

Figure 7. Histograms of the slopes for the SMCI (blue) and PMCI (red) cases. There appears to be two distinct subgroups in the SMCI group. Solid lines show a mixture distribution of two normal curves fitted to the slopes of the SMCI cases. The areas of the histograms are scaled to one. (SD: standard deviation, Q1: 25th quartile, Q3: 75th quartile).

Figure 8. Visualizations for multimodal longitudinal data from three patients. Left panel: Longitudinal DSF showing on the rows the DSI values of individual tests at different time points. Right panel: The dashed line with white circles present regression line of the total DSI values over time for a patient, the black squares show the total DSI values of the patient, the horizontal line indicates a threshold were accuracy of 85% is achieved, and the vertical line shows the current age of the patient.

Figure 9. Observed and predicted annual hippocampal atrophy rates for the models with L1 and L2 norms in the regularization. Analysis 1 included the full set of variables, Analyses 2-4 included only the MRI features.

Figure 10. Kaplan-Meier curves for the optimal set of variables from forward selection. All except sex stratified in tertiles. The curves were plotted using raw data, without imputation.

List of Tables

Table 1. Core biomarkers of Alzheimer's disease

Table 2. Summary of the study populations

Table 3. Characteristics of the population in Study I

Table 4. Characteristics of the population in Study II

Table 5. Characteristics of the population in Study III

Table 6. Characteristics of the population in Study IV

Table 7. Summary of the modelling and performance evaluation methods

Table 8. Classification performance of the classifiers evaluated using cross-validation in the ADC cohort

Table 9. Classification performance of the classifiers evaluated by training with the ADC cohort and testing with the pooled DCN and BAR cohorts

Table 10. Regression parameters of longitudinal DSI values calculated using all variables

Table 11. Classification performance of the regression parameters of the longitudinal DSI values obtained using different data modalities

Table 12. Performance of the models to predict annual hippocampal atrophy rate (models with L1 and L2 norms in the regularization)

Table 13. Performance of the models to predict annual hippocampal atrophy rate, dichotomized to fast vs. slow (models with L1 and L2 norms in the regularization)

Table 14. Selected MRI features when the models were trained with the ADNI 1 cohort and tested with the AIBL cohort (Analysis 4)

Table 15. Associations between baseline variables and mortality

SYMBOLS AND ABBREVIATIONS

Aβ	Amyloid-β
AD	Alzheimer's disease
ADAS-cog	Alzheimer's Disease Assessment Scale-cognitive subscale
ADC	Amsterdam Dementia Cohort
ADNI	Alzheimer's Disease Neuroimaging Initiative
ADNI 1	Alzheimer's Disease Neuroimaging Initiative, phase 1
ADRDA	Alzheimer's Disease and Related Disorders Association
AIBL	Australian Imaging Biomarkers and Lifestyle Study of Ageing
APOE	Apolipoprotein E
AUC	Area under the receiver operating characteristic curve
BACC	Balanced accuracy
BAR	A cohort collected in Barcelona
BLSA	Baltimore Longitudinal Study of Ageing
CDR	Clinical Dementia Rating
CDR-GLOB	Clinical Dementia Rating Global score
CDR-SB	Clinical Dementia Rating Sum of Boxes score
CDSS	Clinical decision support system
CERAD	Consortium to Establish a Registry for Alzheimer's Disease
cGCA	Computed global cortical atrophy score
CI	Confidence interval
cMTA	Computed medial temporal lobe atrophy score
CPH	Cox proportional hazards
CSF	Cerebrospinal fluid
CV	Cross-validation
DCN	German Dementia Competence Network
DLB	Dementia with Lewy bodies
DMSS	Dementia Management Support System
DMSS-R	Revised version of Dementia Management Support System
DSF	Disease State Fingerprint
DSI	Disease State Index

EEG	Electroencephalography
FDG	Fluorodeoxyglucose
FDG-PET	Fluorodeoxyglucose positron emission tomography
FCRT	Free and Cued Selective Reminding Test
FN	Number of false negatives
FP	Number of false positives
FTD	Frontotemporal dementia
GCA	Global cortical atrophy
HC	Healthy control
HR	Hazard ratio
IWG	International Working Group
LAD	Least absolute deviation
LMIC	Low- and middle-income country
LMS	Least mean square
MAE	Mean absolute error
MCI	Mild cognitive impairment
MMSE	Mini-Mental State Examination
MoCA	Montreal Cognitive Assessment
MRI	Magnetic resonance imaging
MSE	Mean squared error
MTA	Medial temporal lobe atrophy
NACC	the National Alzheimer's Coordinating Center
NB	Naïve Bayes
NFT	Neurofibrillary tangle
NIA-AA	National Institute on Aging-Alzheimer's Association
NINCDS	National Institute of Neurological and Communicative Disorders
NPV	Negative predictive value
OASIS	Open Access Series of Imaging Studies
PA	Parietal atrophy
PDD	Parkinson's disease dementia
PDF	Probability density function
PET	Positron emission tomography
PMCI	Progressive mild cognitive impairment
PPV	Positive predictive value
PSCD	Progressive subjective cognitive decline

P-tau	Tau phosphorylated at threonine 181
P-tau217	Tau phosphorylated at threonine 217
RAVLT	Rey Auditory Verbal Learning Test
RF	Random Forests
RMSE	Root mean squared error
ROI	Region of interest
ROC	Receiver operating characteristic
SCD	Subjective cognitive decline
SCD-I	Subjective Cognitive Decline Initiative
SHIMR	Sparse high-order interaction model with rejection option
SMCI	Stable mild cognitive impairment
SSCD	Stable subjective cognitive decline
SVM	Support vector machine
TBM	Tensor-based morphometry
TMT-A	Trail Making Test A
TMT-B	Trail Making Test B
TN	True negatives
TP	True positives
VaD	Vascular dementia
VAT	Visual Association Test
VBM	Voxel-based morphometry
WMH	White matter hyperintensities
Z	Time point zero

1 INTRODUCTION

1.1 Background

Alzheimer's disease (AD) is a neurodegenerative disease, which leads to loss of independence as it progresses towards severe dementia. Due to ageing populations all over the world, prevalence of AD is increasing, and AD is becoming a burden for healthcare systems, economies, individual patients, and their next of kin. Currently, there is no cure for AD and approved treatments only alleviate the symptoms and do not modify the disease process (Cummings et al., 2016). Cost to societies could be reduced considerably if the therapies would delay the onset of the disease or slow down its progression (Brookmeyer et al., 2007; Sloane et al., 2002). Different disease-modifying therapies are being studied (Cummings et al., 2016, 2020) and they should be started as early as possible to be effective (Duara et al., 2009; Galimberti & Scarpini, 2011). In addition to possible drug treatment, lifestyle plays a role in dementia prevention. Targeting multidomain lifestyle interventions to individuals at risk of dementia seems to be an effective way to prevent dementia and decline in cognition (Kivipelto et al., 2018; Ngandu et al., 2015).

However, the early diagnosis of AD is challenging and there are considerable delays in the diagnosis. The average time from initial symptoms to diagnosis has been reported to be one to three years (Bond et al., 2005; Cattel et al., 2000; de Miranda et al., 2011; Speechly et al., 2008). Some of the reasons for the delayed diagnosis are related to patients or their next of kin, these include a belief that memory problems are part of normal ageing, stigma, fear of losing independence and competence, and limited access to healthcare services (Bond et al., 2005; Bradford et al., 2009; Dubois, Padovani, et al., 2016). Whereas others are related to healthcare professionals, these include lack of knowledge in identifying early symptoms of AD, primary care practitioners' lack of confidence in ability to diagnose AD, misdiagnosis and diagnostic uncertainty, and reluctance to disclose the diagnosis when no treatments are available (Bond et al., 2005; Bradford et al., 2009; Dubois, Padovani, et al., 2016). In addition, there are disease related factors hampering the diagnosis: pathological changes in the brain start years before the first clinical symptoms become visible; other dementias and conditions have similar and overlapping symptoms; there is no single test for

AD diagnosis, instead the diagnosis requires large number of different tests, like neuropsychological evaluations and magnetic resonance imaging (MRI). The clinicians must interpret all the different test results to understand the patient's condition and to choose the correct treatment. This excessive amount of data can cause information overload, thus hampering the clinicians' decision making.

Clinical decision support systems (CDSSs) can help the clinicians to manage and interpret large volumes of data. Especially, less experienced clinicians may benefit more from these kinds of tools. Machine learning methods use statistics to search for patterns in data. They compare a patient's measurement values to data from previously diagnosed cases and provide predictions or classifications based on the data. Machine learning and data visualization methods can be part of the CDSSs and support the clinicians in diagnostics so that the interventions could be started at the correct phase and patients would maintain their independence longer. In addition, the machine learning methods can potentially help in identifying suitable patients for clinical trials studying effects of the disease-modifying therapies.

Many traditional and modern machine learning methods are "black boxes", meaning that they take the input data and provide the prediction, but they do not explain why a certain result was given or what variables were the most influential. In medicine, machine learning methods must often be transparent to be accepted by healthcare professionals. One of the transparent methods is Disease State Index (DSI) and its visual counterpart Disease State Fingerprint (DSF), which show both the prediction and contribution of each variable to provide a comprehensive view of a patient's condition. This thesis utilized DSI and other machine learning methods for developing prediction models for different phases of AD, starting from normal cognition and ending to death.

1.2 Outline of the thesis

This thesis is structured as follows. Chapter 2 describes background for this thesis. It starts with an overview of dementia and then focusses on characteristics and diagnostics of AD. It also shows a few examples of CDSSs developed for early diagnosis of AD and reviews the state-of-the-art of the machine learning methods applied to AD progression. Chapter 3 describes the aims of this thesis. Chapter 4 first presents the study populations, then it describes the utilized machine learning and modelling methods, and finally it depicts the used performance evaluation methods.

Chapter 5 presents the main results of the studies. Chapter 6 summarizes accomplishment of the objectives, compares the research contribution to the previous studies, discusses importance of the findings and limitations of the studies, and proposes topics for future research. Finally, Chapter 7 concludes this thesis.

2 REVIEW OF THE LITERATURE

2.1 Dementia

Dementia is a syndrome, a group of symptoms, which can have multiple causes (Alzheimer's Association, 2018). Patients with dementia experience deterioration in their cognitive functions, such as memory, language, learning, and problem solving, which result in difficulties to perform everyday activities (Alzheimer's Association, 2018). Due to ageing populations, dementia is a global problem. Worldwide, 46.8 million people lived with dementia in 2015, and the number is projected to almost double every 20 years, reaching 131.5 million by 2050 (Prince et al., 2015). Much of this increase will be seen in low- and middle-income countries (LMICs). In 2015, 58% of people with dementia lived in LMICs and 22% in Europe, whereas in 2050, 68% live in LMICs and 14% in Europe (Prince et al., 2015). On the societal level, global costs of dementia have been estimated to be US$ 818 billion in 2015, which represents 1.1% of global gross domestic product (Prince et al., 2015). Most of the costs are attributed to the costs of care: 40% direct social care costs (paid and professional home care, residential and nursing home care) and 40% unpaid informal care. Direct medical costs cover only 20% of the total costs (Prince et al., 2015).

AD is the most common cause of dementia, accounting for 60-80% of the cases (Alzheimer's Association, 2018). Other common forms include vascular dementia (VaD), Lewy body dementias, frontotemporal dementia (FTD, also known as frontotemporal lobar degeneration), and mixed dementia. This thesis focuses on AD which is presented in detail in the later chapters. Other dementias are shortly described here.

VaD is the second most common cause of dementia after AD (O'Brien & Thomas, 2015) and attributable to various cerebrovascular pathologies, e.g., infarcts in major or minor cerebral vessels, haemorrhages, or cerebral small vessel disease (Iadecola, 2013; Kalaria, 2018). Because cerebrovascular lesions are heterogeneous, clinical symptoms also vary, depending on location and severity of the lesions (Korczyn et al., 2012; O'Brien & Thomas, 2015). There can be deficits in attention, information processing, and executive function; neuropsychiatric symptoms, aphasia, parkinsonism, and gait abnormalities can also occur (Korczyn et al., 2012;

O'Brien & Thomas, 2015). Due to the heterogeneous nature of VaD, the term vascular cognitive impairment has been taken into use to cover various vascular causes, clinical symptoms, and severities of the condition (Iadecola, 2013; O'Brien et al., 2003).

Lewy body dementias are the second most common type of dementia in people above 65 years and they consists of dementia with Lewy bodies (DLB) and Parkinson's disease dementia (PDD) (Walker et al., 2015). In both conditions, so-called Lewy bodies, aggregates of the protein α-synuclein, accumulate inside neurons (Gomperts, 2016; Jellinger & Korczyn, 2018; Walker et al., 2015). Clinical features of DLB and PDD include cognitive impairment, fluctuations in attention and wakefulness, visual hallucinations, and parkinsonism (Gomperts, 2016; Jellinger & Korczyn, 2018). In DLB, dementia starts before or at the same time with parkinsonism or motor symptoms, whereas in PDD, dementia starts after the motor symptoms related to Parkinson's disease (Gomperts, 2016; Jellinger & Korczyn, 2018; Walker et al., 2015). However, there is a discussion ongoing whether the DLB and PDD are the same or completely distinct diseases (Gomperts, 2016; Jellinger & Korczyn, 2018).

FTD is a heterogeneous group of neurodegenerative disorders, affecting mainly frontal and temporal lobes of the brain; it is also a common cause of dementia in patients under 65 years (Bang et al., 2015; Mann & Snowden, 2017). FTD is divided into three clinical variants: behavioural-variant frontotemporal dementia, associated with early behavioural and executive deficits; primary progressive aphasia, characterized by deficits in speech, grammar, and word output; semantic-variant primary progressive aphasia, causing problems in semantic knowledge and naming (Bang et al., 2015).

Mixed dementia is a condition in which pathologies of two or more dementias occur at the same time (Alzheimer's Association, 2018). The most common combination is AD together with VaD, followed by AD with DLB, and AD with VaD and DLB. The combination of DLB and VaD is not as common.

2.2 Alzheimer's disease

AD was first discovered in 1906 by Alois Alzheimer who found specific pathological changes in the brain of his patient who suffered from progressive cognitive impairment and other mental symptoms (Hippius & Neundörfer, 2003; Maurer et al., 1997). After a vast amount of research, AD is nowadays known to be a progressive,

irreversible, neurodegenerative disease which cannot be cured yet (Alzheimer's Association, 2018). As the disease progresses, patients with AD lose their independence and need more help in activities of daily living than other older people (Alzheimer's Association, 2018). This affects quality of life of the patients as well as families and other informal caregivers. Increased care needs may cause informal caregivers emotional stress and depression; new or worsening health problems; and financial issues (Alzheimer's Association, 2018). In the United States, an estimated 6.2 million people aged 65 years and older lived with Alzheimer's dementia in 2021 and more than 11 million Americans provided unpaid care for people with dementia (Alzheimer's Association, 2021).

2.3 The continuum of Alzheimer's disease

AD is not anymore viewed as discrete clinical stages; rather it is seen as a continuum with certain progressive pathological changes in the brain, related biomarker findings and clinical manifestations, appearing in a sequential and overlapping manner (Aisen et al., 2017). The continuum begins with a long preclinical phase, in which pathological changes start but no clinical symptoms are present yet. It continues to symptomatic phase where progressing pathological changes cause increasingly evident cognitive and functional impairments. Finally, it ends to severe dementia and death.

2.3.1 Pathological changes of the brain in Alzheimer's disease

The aetiology of AD is a complex process and it is not yet fully understood. Nevertheless, it is known that pathological changes in the brain can occur years or even decades before the first clinical symptoms appear (Alzheimer's Association, 2018). The changes include 1) accumulation of *extraneuronal* amyloid-β (Aβ) plaques caused by imbalanced production and clearance of the protein Aβ; 2) accumulation of *intraneuronal* neurofibrillary tangles (NFT), composing mainly of abnormal form of the protein tau, tau phosphorylated at threonine 181 (p-tau); 3) progressing dysfunction and loss of synapses and death of neurons leading to decreased cerebral glucose metabolism and atrophy of the brain (Alzheimer's Association, 2018; Raskin et al., 2015; Serrano-Pozo et al., 2011). Accumulation of different forms of Aβ is thought to impair synaptic activity and cause neuronal dysfunction; NFTs disrupt the transport of nutrients, waste products and other organelles inside the neurons

(Raskin et al., 2015). Other pathological changes include inflammation, oxidative stress, and altered neuronal ionic homeostasis (Alzheimer's Association, 2018; Raskin et al., 2015). These changes in the brain result in neurodegeneration.

2.3.2 Biomarkers of Alzheimer's disease

Biomarkers of AD are associated with the pathological processes of the brain described in the previous section. Table 1 summarizes the most studied and validated biomarkers, including markers measured from cerebrospinal fluid (CSF) samples, positron emission tomography (PET), and MRI. These biomarkers can be grouped into markers of Aβ accumulation, markers of tau pathology, and markers of neurodegeneration (Jack et al., 2016, 2018). Not all these biomarkers are sensitive to AD, and they can be present in other conditions as well. Vast amount of research has been conducted to find other potential biomarkers of AD from blood, CSF, genetics, and functional MRI (Dennis & Thompson, 2014; Gustaw-Rothenberg et al., 2010; Hampel et al., 2008; Henriksen et al., 2014; Lehmann et al., 2013; Lewczuk et al., 2018; Shaw et al., 2007; R. Sperling & Johnson, 2013; Wolfsgruber et al., 2017).

Table 1. Core biomarkers of Alzheimer's disease

Biomarker	Pathology	Finding in AD
CSF		
Aβ	Aβ deposition	- Decreased concentration
P-tau	Tau pathology	- Increased concentration - Associated with NFT burden - Fairly specific for AD
Total tau	Neurodegeneration	- Increased concentration - Elevated also in ischaemic and traumatic brain injury, Creutzfeld-Jakob disease, and stroke
PET scan		
Amyloid PET	Aβ deposition	- Increased retention of amyloid tracers (^{11}C-Pittsburgh Compound B, ^{18}F ligands, such as flutemetamol, florbetaben, florbetabir) - The tracers bind to fibrillary, and not to soluble Aβ or diffuse plaques.
Tau PET	Tau pathology	- Increased retention of tau tracer (flortaucipir is a first-generation tracer; others in development) - New biomarker
FDG-PET	Neurodegeneration	- Reduced retention of ^{18}F-fluorodeoxyglucose (FDG) - Measure of reduced glucose metabolism in the brain and a marker of synaptic dysfunction
Structural MRI	Neurodegeneration	- Decreased volume of the brain structures - Atrophy of temporal lobe structures, especially hippocampus and entorhinal cortex - Volume of hippocampus decreases also in normal ageing, several other neurodegenerative and non-neurodegenerative diseases (e.g., diabetes, sleep apnoea, bipolar disorder)

Based on (Aisen et al., 2017; Hampel et al., 2008; Jack et al., 2010, 2016, 2018; R. Sperling & Johnson, 2013).

2.3.3 Risk factors of Alzheimer's disease

There are several factors modifying the risk of developing cognitive decline, AD, or other dementias. The most important risk factors of AD are older age, a family history of AD, and carrying apolipoprotein (APOE) ε4 gene (Alzheimer's Association, 2018). Carrying APOE ε2 gene, instead of ε4, is protective against AD (Alzheimer's

Association, 2018). Brain health is also affected by health of heart and blood vessels because the brain consumes considerable amount of oxygen. Thus, healthy diet, regular physical activity, and proper management of cardiovascular risk factors, such as diabetes, obesity, smoking, and hypertension, are modifiable factors that can reduce the risk of cognitive decline and dementia (Alzheimer's Association, 2018). Cognitive reserve is another factor, which partly explains why some people cope better than others with a certain level of pathological load in the brain (Stern, 2009). Cognitive reserve refers to inter-individual differences in cognitive processing or neural networks in the healthy brain and to ability of the brain to use alternative networks instead of damaged ones. Longer education, occupational attainment, and active participation in leisure time activities increase cognitive reserve and, thus, reduce risk of dementia (Stern, 2012). Higher cognitive reserve may delay the onset of clinical symptoms of AD, but after the onset, cognitive decline can be more rapid because underlying pathologies are more severe than in individuals with lower cognitive reserve (Stern, 2012). Finally, moderate or severe traumatic brain injury increases the risk of dementias and the risk may be even higher in people who experience repetitive traumatic brain injuries (Alzheimer's Association, 2018).

2.3.4 Temporal evolution of the biomarkers in Alzheimer's disease

Findings in AD biomarker research led to a development of a hypothetical model describing temporal ordering and evolution of the major AD biomarkers (Jack et al., 2010). A few years after its publication, the model was updated on the basis of gained knowledge (Jack et al., 2013). The updated model is depicted in Figure 1. According to it, the biomarkers become abnormal in a temporally ordered manner as the disease progresses, and their rates of change follow a non-linear sigmoidal time course. The biomarkers of Aβ deposits become abnormal first and have largely plateaued by the time of clinical symptoms appear. They are followed by the biomarkers of tau-mediated neuronal injury and dysfunction, correlating with severity of the clinical symptoms. The last biomarkers to become abnormal are uptake of fluorodeoxyglucose on PET imaging (FDG-PET) and structural MRI. Accumulation of Aβ plaques is seen necessary but not sufficient, to produce clinical symptoms of AD. In the model, cognitive impairment is presented as a zone bounded by limits of low and high risk. This emphasizes that individuals respond to pathophysiological load of AD uniquely,

depending on their genetic factors, cognitive reserve, lifestyle, other brain pathologies, and comorbidities. Thus, a high-risk individual can suffer from more severe cognitive impairment than a low-risk individual with a similar biomarker profile.

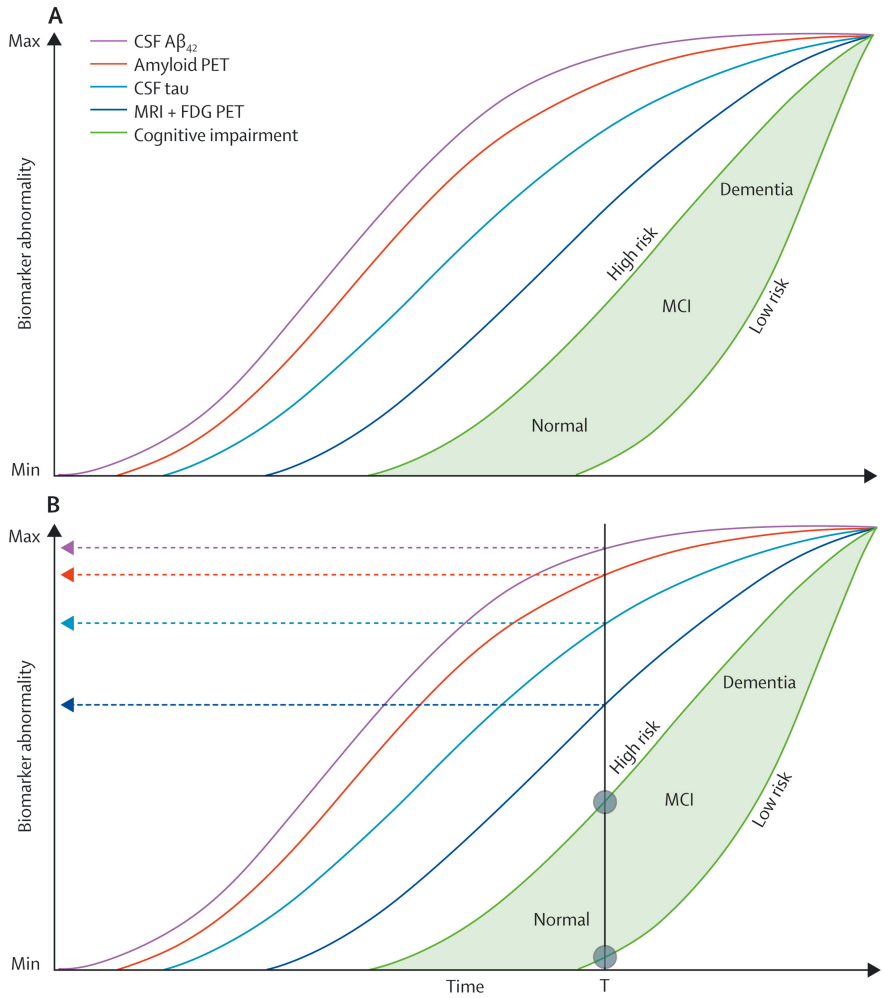


Figure 1. Panels A and B: A hypothetical model describing temporal evolution of the major Alzheimer's disease biomarkers and cognitive impairment which is shown as a light green area bounded by limits for high and low risk. Panel B: Operational use of the model. The black vertical line denotes a given time (T) and the horizontal dashed arrows show values of each biomarker at time T. The grey circles highlight that patients with similar biomarker profiles at time T can have different levels of cognitive impairment, depending on their risk factors for AD (Jack et al., 2013)[1].

[1] Reprinted from The Lancet Neurology, 12(2), Jack, C. R., Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., Shaw, L. M., Vemuri, P., Wiste, H. J., Weigand, S. D., Lesnick, T. G., Pankratz, V. S., Donohue, M. C., & Trojanowski, J. Q., Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers, 207-216, Copyright (2013), with permission from Elsevier.

2.3.5 Diagnostic guidelines for Alzheimer's disease

The first diagnostic guidelines for AD were developed in 1984 by the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS) and the Alzheimer's Disease and Related Disorders Association (ADRDA) (G. McKhann et al., 1984). These guidelines have been updated several times by different work groups as the knowledge about the biomarkers of AD has evolved. It should be noted that most of the updated guidelines were intended only for research purposes and not for diagnosis in a routine clinical setting. This section describes similarities, differences, and nomenclatures of the different guidelines.

The first guidelines by NINCDS-ADRDA covered only the dementia phase of AD and were based only on a clinical assessment and laboratory tests were used only for exclusion of other conditions (G. McKhann et al., 1984). Diagnosis of ***probable AD*** required a typical insidious onset of dementia with progression and without any other explaining systemic or brain diseases. Diagnosis of ***possible AD*** was used if other diseases were present and AD was considered as the most likely cause of dementia or if the presentation or course of dementia was aberrant. Diagnosis of ***definite AD*** required postmortem histopathological samples.

In 2011, the National Institute on Aging-Alzheimer's Association (NIA-AA) developed separate diagnostic guidelines for preclinical AD, mild cognitive impairment (MCI), and AD dementia (Albert et al., 2011; Jack et al., 2011; G. M. McKhann et al., 2011; R. A. Sperling et al., 2011). In addition to the clinical assessment, these guidelines also incorporated the biomarkers of Aβ accumulation (amyloid PET, Aβ in CSF) and neurodegeneration (total tau and p-tau in CSF, FDG-PET, MRI). ***Preclinical AD*** is the phase were pathophysiological process of AD has begun in the brain but there are no or only subtle clinical symptoms that do not yet fulfil the criteria for MCI or AD dementia (R. A. Sperling et al., 2011). The diagnosis of preclinical AD is based on abnormal findings in biomarkers of Aβ accumulation (Stage 1); Aβ accumulation and neurodegeneration (Stage 2); Aβ accumulation, neurodegeneration, and subtle cognitive decline (Stage 3). ***MCI*** is a syndrome in which cognitive decline is greater than expected for an individual's age and education but does not interfere with performance of daily activities (Gauthier et al., 2006; Petersen, 2009, 2016). MCI with memory complaints (amnestic MCI) is a risk factor for AD dementia, however, there are also other causes for MCI and not all patients with MCI will progress to AD dementia (Gauthier et al., 2006; Petersen, 2009, 2016). The NIA-AA guidelines used the term ***MCI due to AD*** to refer to the symptomatic predementia phase of AD (Albert et al., 2011). Similar to the NINCDS-ADRDA

guidelines, NIA-AA divided ***dementia due to AD*** to probable and possible (G. M. McKhann et al., 2011). In the diagnosis of MCI or dementia due to AD, the clinical assessment forms the core criteria and the biomarkers provide additional information for definition of certainty (unlike, intermediate, high) that the underlying pathology is AD (Albert et al., 2011; Jack et al., 2011; G. M. McKhann et al., 2011). If the biomarkers are conflicting, indeterminate, or unavailable, the diagnosis is only based on the clinical assessment (Albert et al., 2011; Jack et al., 2011; G. M. McKhann et al., 2011).

International Working Group (IWG) is another group that has updated diagnostic guidelines for AD several times (Dubois et al., 2007, 2010, 2014; Dubois, Hampel, et al., 2016). In their most recent guidelines, the diagnosis required evidence from both clinical assessment and biomarkers (Dubois et al., 2014). The biomarker evidence of AD included 1) amyloid PET or 2) Aβ and total tau or p-tau in CSF or 3) presence of AD autosomal dominant mutation. FDG-PET or MRI were excluded because they might better serve in monitoring the progression of the disease. IWG has also recently updated their diagnostic guidelines for preclinical AD (Dubois, Hampel, et al., 2016). The diagnosis ***preclinical AD*** is given when the biomarkers of both Aβ (CSF Aβ or amyloid PET) and tau (CSF or PET tau) are beyond threshold values in asymptomatic individuals. The diagnosis ***asymptomatic at risk for AD*** is achieved when only one of the Aβ or tau are abnormal in cognitively normal individuals. In addition, IWG has proposed a lexicon for AD research (Dubois et al., 2010). According to it, ***prodromal AD*** or ***predementia stage of AD*** is a similar concept as MCI due to AD in the NIA-AA guidelines and the term MCI refers to cases who have cognitive impairment but do not fulfil the criteria for prodromal AD, i.e., do not have memory symptoms or biomarkers specific for AD. In the lexicon, AD dementia was divided into ***typical AD*** (the most common clinical phenotype), ***atypical AD*** (less common and characterized clinical phenotype) and ***mixed AD*** (typical AD and evidence of other comorbidities).

The last concept presented here is ***subjective cognitive decline*** (SCD) which is related to many different conditions (e.g. normal ageing, psychiatric and neurologic disorders, substance and medication use) (Jessen et al., 2014). Research criteria for SCD in relation to AD was proposed by Subjective Cognitive Decline Initiative (SCD-I) (Jessen et al., 2014). They suggested that SCD occurs at the late stage of preclinical AD, just before the progression to MCI/prodromal AD. The criteria for SCD requires 1) self-experienced persistent decline in cognitive capacity in comparison to a previously normal status and unrelated to an acute event and 2) normal age-, gender-, education-adjusted performance on standardized cognitive tests for

MCI/prodromal AD (Jessen et al., 2014). SCD-I presented also a ***SCD plus*** criteria, which lists features increasing the likelihood of preclinical AD in subjects with SCD. These features include subjective decline in memory, onset of SCD within the last five years, age at onset ≥60 years, worries associated with SCD, feelings of worse performance than others at the same age, confirmation of cognitive decline by an informant, presence of APOE ε4, and biomarker evidence of AD (Jessen et al., 2014).

The most recent guidelines for research published by NIA-AA in 2018 are based on an initial suggestion by Jack et al. (Jack et al., 2016) and they are a complete opposite to the first NINCDS-ADRDA guidelines (Jack et al., 2018). The guidelines are based only on detection of underlying AD pathology either by postmortem samples or in vivo by biomarkers. Clinical symptoms are not part of the diagnosis at all. NIA-AA suggested to group biomarkers into those of Aβ deposition, pathologic tau, and neurodegeneration, forming an AT(N) classification system. Group A biomarkers include CSF $A\beta_{42}$, CSF $A\beta_{42}/A\beta_{40}$ ratio, and amyloid PET; Group T biomarkers include CSF p-tau or tau PET; Group (N) biomarkers include volumetric MRI, FDG-PET, and CSF total tau. These biomarker groups are dichotomized to normal(-)/abnormal(+) using threshold values. This results in eight different biomarker profiles (A-T-(N)-, A+T-(N)-, A+T+(N)-, etc.). The biomarker profile is normal if all A, T and (N) are negative. Alzheimer's pathologic change is found when only A is positive. Alzheimer's disease is present if both A and T are positive. Profile A+T-(N)+ corresponds to Alzheimer's and concomitant suspected non-Azheimer's pathologic change and, finally, negative A refers to non-AD pathologic change. N is in the parenthesis because biomarkers of neurodegeneration are not specific to AD. NIA-AA also proposed a separate staging for cognitive symptoms (unimpaired/MCI/dementia) because cognitive staging is part of many ongoing studies and it has been adopted by many medical practitioners. Thus, each individual has both a biomarker profile and a cognitive stage.

2.3.6 Risk factors for mortality in Alzheimer's disease

Dementia due to AD shortens the life expectancy of an individual, average survival time being 4 to 8 years after the diagnosis in patients aged 65 years and older (Alzheimer's Association, 2018). However, the survival time varies considerably between individuals, some living as long as 20 years after the diagnosis. Several studies have investigated risk factors for mortality in AD. Older age and male gender have been found to associate strongly with mortality in dementia and AD (Todd et al.,

2013). Studies assessing impact of disease severity have used several different scales and instruments to measure global, cognitive, and functional impairment. Global impairment measured with Clinical Dementia Rating (CDR) was associated with mortality, and studies on cognitive and functional impairment have provided contradicting results (Todd et al., 2013). Diabetes mellitus, smoking, coronary heart disease, and congestive heart failure were associated with mortality in dementia, but meta-analysed studies included only older patients (mean age 74-87 years) (van de Vorst et al., 2016). A few studies have also examined associations between AD biomarkers and mortality. A study on memory clinic patients found microbleeds to be strongly associated with mortality in AD and in the whole study population (patients with SCD, MCI, AD, other dementia, or other diagnosis) (Henneman, Sluimer, et al., 2009). White matter hyperintensities were associated to a lesser extent with mortality in the whole population, but not in AD (Henneman, Sluimer, et al., 2009). Global cortical atrophy was associated with mortality only in the younger patients (<68 years) (Henneman, Sluimer, et al., 2009). Another study found frontal and medial temporal atrophy to be associated with mortality in AD (Nägga et al., 2014). Studies on CSF biomarkers and mortality have reported mixed results. A recent study found CSF $A\beta_{42}$ to be associated with mortality in patients with AD having abnormal values for all three CSF biomarkers ($A\beta_{42}$, total tau, p-tau), corresponding to the recent AT(N) classification system (Boumenir et al., 2019). Another study found no associations between CSF biomarkers ($A\beta_{42}$, total tau, p-tau) and mortality in patients diagnosed with probable or possible AD, i.e., not considering AT(N) classification (Nägga et al., 2014). Instead, it found cerebral inflammation to be associated with mortality. Wallin and colleagues (Wallin et al., 2010) used k-means clustering to group patients based on their CSF biomarkers ($A\beta_{42}$, total tau, p-tau). They found that a group with extreme CSF biomarker values had a higher mortality. However, this subgroup of patients was small (n=12, 8% of the study population). Degerman Gunnarson and co-workers (Degerman Gunnarsson et al., 2014) found that having CSF total tau in the highest quartile increased the risk of dying in severe dementia.

2.4 Clinical decision support systems for dementia

CDSSs can be defined as "*software that designed to be a direct aid to clinical decision-making, in which the characteristics of an individual patient are matched to a computerized clinical knowledge base and patient-specific assessments or recommendations are then presented to the clinician or the patient for a decision*" (Sim et al., 2001). Common tasks in which the CDSSs may aid

include diagnostics, assessment of risk for developing a disease or adverse event, prediction of treatment outcomes or hospital resource needs, planning of interventions (e.g. treatment or surgery plans), monitoring a patient's state over time, medication prescription (e.g. drug interactions, cheaper medications), providing alarms, following clinical guidelines, and many more (R. T. Sutton et al., 2020; Tohka & van Gils, 2021). The CDSSs are commonly divided into knowledge-based or non-knowledge-based systems. The knowledge-based systems include a rule base, often in the form of IF-THEN statements, and the rules can be created using literature-based, practice-based, or patient-directed evidence (R. T. Sutton et al., 2020). The nonknowledge-based systems utilize data-driven approaches, such as statistical pattern recognition, machine learning, or artificial intelligence, which infer decisions from the previously measured data (R. T. Sutton et al., 2020). Following paragraphs present three examples of CDSSs developed for dementia.

PredictAD software tool was designed for early diagnosis of AD. It collects relevant data measured from the patients in one place, quantifies automatically MRIs, visualizes heterogeneous data in an understandable way, and includes machine learning methods for combining data from several data modalities and predicting AD progression (Antila et al., 2013; Mattila et al., 2012; Soininen et al., 2012). The PredictAD tool has later been expanded to differential diagnosis of dementia (Koikkalainen et al., 2016; Tolonen et al., 2018). Studies I and II in this thesis are closely related to this tool.

Another example is a prototype CDSS which evaluates severity of AD on a continuous scale (Bucholc et al., 2019). The severity score is based on Kernel Ridge Regression of total scores from four cognitive and functional assessments, namely Functional Assessment Questionnaire, Alzheimer's Disease Assessment Scale 13, Montreal Cognitive Assessment, and Mini-Mental State Examination. This CDSS presents a patient profile with relevant medical information and visualizes the severity score and the total scores of the four assessments. As it does not require inputs from costly biomarkers, like MRI, it can be implemented in primary care.

The last example is Dementia Management Support System (DMSS) and its revised version (DMSS-R) which assist healthcare professionals in examining patients with suspected dementia (Lindgren, 2008, 2011; Lindgren et al., 2002). DMSS is based on a set of clinical practice guidelines for dementia care. It provides interactive support through the diagnostic process, reminds and alerts in the case of missing or questionable data, and proposes suitable interventions and possible diagnoses. The revised version also provides more extensive support for diagnostic hypothesis generation and evaluation.

2.5 Development of machine learning systems

Machine learning is a subcategory of artificial intelligence and it aims to teach computers to solve problems by learning from past experiences (Goodfellow et al., 2016). This means that computers use statistics to search for patterns in data, instead of using fixed and hard-coded programs. Machine learning can be divided into supervised, unsupervised, and reinforcement learning. In supervised learning, both input and output variables are known and used for training the model, i.e., finding the mapping between the input and output variables. Supervised learning can further be divided into classification, where the output variable is categorical, and regression, where the output variable is continuous. In unsupervised learning, the output variable is unknown or is not available. Unsupervised learning algorithms aim to find clusters or groupings in data. They can also be used for dimension reduction by searching for the most important components that best describe the original data without losing too much of information (e.g., principal component analysis). Reinforcement learning is close to human learning which is based on interacting with an environment and analysing consequences of performed actions (R. S. Sutton & Barto, 2018). In reinforcement learning, an agent has a certain goal which it tries to achieve by taking actions to influence its environment (R. S. Sutton & Barto, 2018). For each action, the environment sends a reward or punishment to the agent, whose only goal is to maximize the total reward in the long run (R. S. Sutton & Barto, 2018). An example of reinforcement learning is AlphaZero algorithm which defeated state-of-the-art algorithms in the games of chess, Go, and shogi (Silver et al., 2018). This thesis will focus on supervised learning.

2.5.1 Design cycle of machine learning systems

A general design cycle of a machine learning system is shown in Figure 2. Not all steps are present in all applications, some steps may overlap so that it is difficult to differentiate them, and some methods combine several steps by their design (e.g., deep learning methods can combine feature extraction, feature selection, and modelling). The steps are iteratively repeated during the development process because different steps will give new insights to data and the problem at hand.

The first step, problem definition, defines aim of the system and things that need to be considered when designing such a system. Understanding of the problem usually requires domain knowledge from the experts in the area, e.g., healthcare professionals.

The second step, data collection, is an essential step in the development. Data collection is often time consuming and resource intensive. Data can be collected from existing databases, e.g., electronic health records, population records, publicly available databases. If suitable data do not yet exist, a data collection study needs to be planned and carried out for collecting the needed input and output variables.

The third step, data pre-processing, assures that data are ready for the later steps. This step may include transforming data sets collected from different sources into a common format, data synchronisation, harmonization of variables, and removal of artefacts and noise.

In the fourth step, data exploration, summary statistics and distributions of variables are studied to get an initial understanding of the data.

In the fifth step, feature extraction, new variables (or features) are calculated from images, signals, or other existing variables. Examples of new variables (features) are volumes of different brain structures calculated from MRI or average heart rate calculated from an electrocardiograph.

In the sixth step, feature selection, the most relevant features are selected as inputs for the classifier. A good feature is simple to extract, invariant to irrelevant changes, insensitive to noise, and discriminates well patterns into categories (Duda et al., 2001). Feature selection can be based on both prior knowledge and empirical data.

The seventh and eighth step, model selection and model training, are closely related. Model training means teaching the model to find patterns in data, i.e., parameters of the model are determined using the training data. Thus, available training data and the problem to be solved affect which type of model will be selected.

The last step, model evaluation, assesses performance of the model. In the evaluation, it is important to use an independent data set which was not utilized in the model training. If the model is trained and tested using the same data, there is a high risk that the model will overfit. This means that the model performs very well with the training data, but it performs poorly on new unseen cases. More complex models are more prone to overfitting than simpler models. Different performance evaluation methods are described in the next chapter.

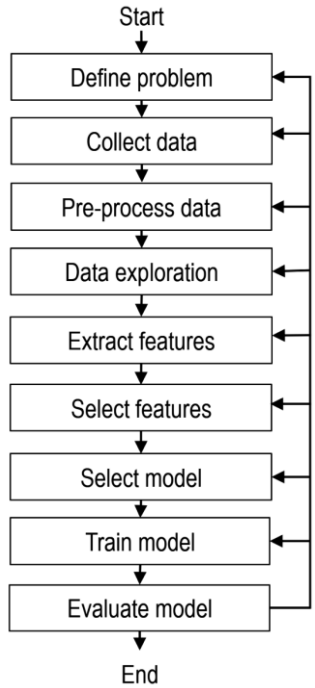


Figure 2. A general design cycle of a machine learning system (based on Duda et al. (2001) and Webb and Copsey (2011))

2.5.2 Performance evaluation of machine learning systems

There are various methods for evaluating performance of the model. In holdout validation, a data set is randomly partitioned into two complementary subsets: one set is used for training the model and the other is used for testing. In k-fold cross-validation (CV), the data set is randomly partitioned into k complementary subsets. The models are trained using the k-1 subsets and tested using the remaining subset not used in the training. Training and testing are repeated k times so that each subset is used k-1 times for training and once for testing. After running all folds, performance metrics are then averaged over the folds. The holdout and cross-validation procedures can be repeated (iterated) several times to obtain robust estimates for the performance metrics. Partitioning into the train and test sets can be stratified, meaning that each fold contains roughly equal proportions of different classes. The best way to estimate generalizability of the models is to train them using one cohort and

test them using another completely different cohort collected, e.g., in a different centre. This approach estimates how well the models generalize and perform in a different setting, e.g., in a different country or slightly different patient group.

Nested CV is used for tuning hyperparameters of the model. In nested CV, the whole data set is first divided into cross-validation folds (outer CV), then the training set of a fold is further divided into cross-validation folds (inner CV). The set of possible hyperparameter values are evaluated in the inner CV and the values providing the best results are selected. Then the model with the selected hyperparameter values is applied to the whole training set of the outer CV and evaluated with the test set of the outer CV. Like in the regular CV, this procedure is repeated for each fold.

2.5.3 Performance metrics used in the evaluation

In classification tasks, common performance metrics are calculated from a confusion matrix which tabulates number of true positives (correctly classified positive cases), true negatives (correctly classified negative cases), false positives (negative cases misclassified as positive), and false negatives (positive case misclassified as negative) (Figure 3). Positive cases refer to diseased subjects and negative cases refer to healthy subjects. The performance metrics include accuracy (the number of correctly classified cases divided by the total number of cases), sensitivity or recall (number of correctly classified positive cases divided by the total number of positive cases), specificity (the number of correctly classified negative cases divided by the total number of negative cases), positive predictive value or precision (PPV, the number of correctly classified positive cases divided by the total number of cases classified as positive), and negative predictive value (NPV, the number of correctly classified negative cases divided by the total number of cases classified as negative). The accuracy is not an optimal measure for performance because it can provide misleading results on imbalanced data sets in which number of cases in each class differs greatly. Balanced accuracy (BACC), defined as an average of sensitivity and specificity, is a better overall performance measure because it takes the uneven class distribution into account. PPV and NPV are interesting measures for clinicians and patients, because they answer the question "what is the chance to really have a disease when the classification result indicates the disease?" However, these measures are also affected by prevalence of the disease (the class distributions). When prevalence increases, PPV increases as well, whereas NPV decreases. This should be taken into account when

interpreting and comparing accuracy, NPV, and PPV of classifiers tested with different datasets. Sensitivity and specificity do not depend on prevalence. All these measures are between 0 and 1 (or 0% and 100%), higher values indicating better performance.

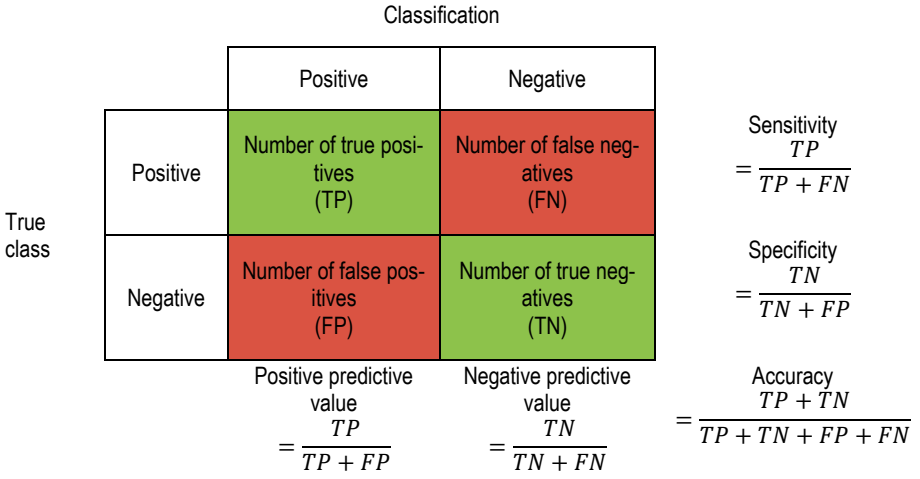


Figure 3. Confusion matrix

Classifiers often produce a probability as an output and above-mentioned performance metrics are calculated by defining a threshold for probability values to classify cases in one of the classes. In a two-class task, sensitivity and specificity values can be calculated for different threshold values. A receiver operating characteristic (ROC) curve can then be created by plotting sensitivity against 1-specificity values. The ROC curve illustrates overall performance of a classifier over different thresholds. A classifier, of which ROC curve goes closest to the point (0, 1), has the best performance. If the ROC curve is close to the diagonal line from the left bottom corner to the right top corner, performance of the classifier is as good as a random guess. The ROC curve can be used for finding an optimal threshold for the classification. Optimality depends on whether the application needs to favour high sensitivity or high specificity. The ROC curve can be summarized into one value, namely area under the ROC curve (AUC). AUC is an additional measure for the overall classification performance. AUC of one means a perfect classification and 0.5 means random classification. As calculation of AUC is based on the ROC curve, it summarizes performance of the classifier over different classification thresholds. Whereas, accuracy and other performance metrics calculated from the confusion matrix, are related to one selected point on the ROC curve. In addition, AUC is not sensitive to imbalanced class distributions.

In regression tasks, performance of the model can be visualized by plotting the predicted values against the observed values. Performance can also be numerically summarized with correlation coefficient (Pearson or Spearman), root mean squared error (RMSE), mean squared error (MSE), and mean absolute error (MAE). Correlation coefficients are between -1 and 1. Values close to 0 indicate no correlation and values close to -1 and 1 indicate strong correlation. Values close to -1 indicate strong opposite correlation (e.g., observed values increase and predicted values decrease). The correlation coefficients are problematic in performance evaluation because they do not reveal the actual differences between the predicted and observed values. E.g., correlation coefficient is one if all predicted values are systematically five points higher than observed values. Thus, RMSE, MSE, and MAE are better performance measures for regression tasks. As RMSE and MSE square the differences between the observed and predicted values, bigger errors have more impact on the result than smaller errors. MAE uses the absolute differences instead of the squared differences; thus, bigger errors are not emphasized, and MAE is more robust against outliers. RMSE and MAE have the same unit as the outcome variable, and MSE has the squared unit, thus, models for different outcomes cannot be compared directly using these measures. The correlation coefficients do not depend on the units of outcomes.

2.6 Machine learning in Alzheimer's disease

2.6.1 Typical research questions

Machine learning has extensively been applied to AD during the past years. Typical research foci in this field include 1) diagnosis, i.e., classification of subjects to different diagnostic classes (e.g., healthy controls (HC) vs. AD, HC vs. MCI, MCI vs. AD, HC vs. MCI vs. AD); 2) prediction of conversion from one disease stage to another, conversion from MCI to AD being the most common; 3) differential diagnosis of dementia; 4) identification of risk factors; and 5) prediction of AD progression using continuous cognitive scores as outcomes (Battista et al., 2020; Bratić et al., 2018; Fabrizio et al., 2021; Kumar et al., 2021; Rathore et al., 2017; Weiner et al., 2015, 2017; R. Zhang et al., 2017). In addition, machine learning has been utilized for mining AD literature; evaluating clinical care and use of care resources of patients with

AD; and investigating relationship between cognition, cognitive reserve, and AD (R. Zhang et al., 2017).

2.6.2 Data sources and data modalities

Several public and private data sources in different sizes have been utilized for machine learning in AD. According to a recent review, most of the studies included at most 1000 subjects (n=34), 15 studies included 1001-10000 subjects, and 12 studies included more than 10000 subjects (Kumar et al., 2021).

The Alzheimer's Disease Neuroimaging Initiative database (ADNI) is the most used publicly available data set (Kumar et al., 2021; Martí-Juan et al., 2020; R. Zhang et al., 2017). The first phase of the ADNI (ADNI 1) was a longitudinal six-year study aiming at testing whether a combination of clinical, cognitive, imaging, genetic, and biochemical biomarkers can measure progression of MCI and AD (Weiner et al., 2010). ADNI 1 aimed to recruit approximately 800 subjects (200 HC, 400 MCI, 200 early AD) of age 55 to 90 years at around 50 sites in the USA and Canada (Petersen et al., 2010; Weiner et al., 2017). ADNI 1 was followed by ADNI GO, ADNI 2, and ADNI 3 studies, which recruited new subjects and continued to follow some of the subjects from the preceding ADNI studies. ADNI has several data modalities measured longitudinally, long follow-up times, a well-organized structure, and an easy access, which together have promoted its wide use. Other utilized public data sets include, e.g., the Australian Imaging Biomarkers and Lifestyle Flagship Study of Ageing (AIBL), a cross-European AddNeuroMed and its curated version ANMerge, the Open Access Series of Imaging Studies (OASIS), the National Alzheimer's Coordinating Center (NACC), and Baltimore Longitudinal Study of Aging (BLSA) (Bratić et al., 2018; Kumar et al., 2021; Martí-Juan et al., 2020; D. Zhang et al., 2011).

Typically, machine learning studies in AD have included data from the core biomarkers of AD and traditional clinical assessments: neuroimaging (MRI, PET, diffusion tensor imaging), neuropsychological and cognitive assessments (e.g. Mini-Mental State Examination (MMSE), Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-cog)), genetics (e.g. APOE, family history), laboratory (e.g. CSF, blood plasma proteins), and demographics (Battista et al., 2020; Bratić et al., 2018; Kumar et al., 2021; Martí-Juan et al., 2020; Rathore et al., 2017). MRI alone or together with other data modalities is one of the most studied data modality (Martí-Juan et al., 2020) as it is a non-invasive method to obtain good quality images of the

brain structures that are affected already in the early phase of AD. Other data modalities have been studied less and typically with limited sample sizes. These modalities include clinical notes, like discharge summaries (Kumar et al., 2021), electroencephalography (EEG) (Graham et al., 2020; Tăuţan et al., 2021), handwriting (Graham et al., 2020; Impedovo & Pirlo, 2018), inertial and other sensors for gait and movement analysis (Cavedoni et al., 2020; Graham et al., 2020; Muurling et al., 2020), voice and speech recordings (Graham et al., 2020; Tăuţan et al., 2021), web-based cognitive testing (Meester et al., 2020), and eye movements (Tăuţan et al., 2021). Aim of the studies on voice, eye movements, gait/movement, web-based cognitive tests, and handwriting was to find non-invasive, low-cost, and easy-to-perform methods for early detection of AD. Another emerging field of study is applying machine learning to omics data obtained from metabolomics, genomics, proteomics, and transcriptomics. These new biomarkers may aid in the diagnosis of AD and reveal underlying factors affecting development and progression of AD (Tan et al., 2021). This thesis will focus on the traditional data modalities. The emerging data modalities and their analyses are not in the scope of this review.

2.6.3 Developed models

As AD has been a popular application area for machine learning in recent years, plethora of different machine learning methods have been applied to AD to solve the research questions described in Chapter 2.6.1. Among the classification studies, support vector machine (SVM) classifier has been one of the most popular machine learning method, especially in the neuroimaging studies (Battista et al., 2020; Martí-Juan et al., 2020; Rathore et al., 2017; Tăuţan et al., 2021). Other used classification methods include decision trees (e.g. simple decision trees, Random Forests, AdaBoost); Bayesian networks (e.g. Naïve Bayes, Bayesian belief networks); neural networks and deep learning (e.g. multilayer perceptron, convolutional and recurrent neural networks, autoencoders, restricted Boltzmann machines); logistic regression with or without regularization; DSI; and linear and quadratic discriminant analysis (Kumar et al., 2021; Martí-Juan et al., 2020; Tăuţan et al., 2021; Weiner et al., 2015, 2017; R. Zhang et al., 2017).

The studies predicting continuous cognitive scores are not as common as classification studies, nevertheless, they have utilized variety of different methods as well. Also in these studies, SVMs are among the most popular methods (Bucholc et al., 2019; Huang et al., 2016; D. Zhang et al., 2012b; Zhu et al., 2016). Several studies

have developed or applied more complex feature extraction and prediction methods which consider multi-modal or longitudinal data. These include kernel ridge regression (Bucholc et al., 2019), neural networks and deep learning (radial basis function and convolutional neural networks) (Dong et al., 2020; Zhu et al., 2016), multi-task learning (D. Zhang et al., 2012a; Zhou et al., 2013), and Random Forests based non-linear sparse regression (Huang et al., 2016). Some studies compared these methods to more traditional regression methods like linear (Zhu et al., 2016), lasso (Dong et al., 2020; Huang et al., 2016; Zhou et al., 2013), ridge (Dong et al., 2020; Huang et al., 2016; Zhou et al., 2013), k-nearest neighbour (Bucholc et al., 2019), decision tree (Zhu et al., 2016), and Gaussian process (Zhu et al., 2016).

Risk factors for AD and its progression have commonly been identified using Cox proportional hazards model (Fabrizio et al., 2021; Golriz Khatami et al., 2020; R. Zhang et al., 2017).

2.6.4 Performance of the models

Direct comparison of performances of the developed models is difficult due to differences in the selected data sets, techniques used for handling missing values, used validation and performance evaluation methods, and reported performance metrics (Martí-Juan et al., 2020; R. Zhang et al., 2017). Most studies have used cross-validation in the performance evaluation. Some studies have also assessed generalizability of the models using an independent validation set (Martí-Juan et al., 2020).

2.6.4.1 Diagnosis of AD, MCI, and conversion from MCI to AD

Main findings in this chapter were extracted from three review articles summarizing recent ADNI publications (Weiner et al., 2017) and machine learning results from neuropsychological (Battista et al., 2020) and neuroimaging data (Martí-Juan et al., 2020). When AD has progressed to the dementia phase, cognitive symptoms and damage to the brain is already evident. Thus, differentiating HCs from patients with Alzheimer's dementia is the easiest classification task for machine learning. Studies classifying HC and AD dementia using only neuropsychological data reported accuracies ranging from 72% to 100% (sensitivity 73-100%, specificity 77-100%, AUC 0.79-0.98) (Battista et al., 2020). Studies using only MRI reported accuracies varying from 83% to 97% and studies using only FDG-PET reported accuracies from 78% to 93% (Martí-Juan et al., 2020; Weiner et al., 2017).

Other classification tasks are harder due to overlapping symptoms and biomarker values between the classes. Studies differentiating HC from MCI using only neuropsychological data reported accuracies from 60% to 98% (sensitivity 45-97%, specificity 67-100%, AUC 0.63-0.99) (Battista et al., 2020). Studies using only MRI reported accuracies varying from 54% to 97% and studies using only FDG-PET reported accuracies from 63% to 70% (Martí-Juan et al., 2020; Weiner et al., 2017). The hardest task is to predict which of the MCI cases will progress to Alzheimer's dementia. This task is also limited by the availability of longitudinal data sets having long enough follow-up periods to detect conversions. Studies using only neuropsychological data reported accuracies from 61% to 85% (sensitivity 50-91%, specificity 48-91%, AUC 0.67-0.93) (Battista et al., 2020). Studies using only MRI reported accuracies varying from 62% to 92% and studies using only FDG-PET reported accuracies from 63% to 81% (Martí-Juan et al., 2020; Weiner et al., 2017).

Different data modalities provide complementary information, thus, several studies developed models combining different modalities. These multimodal models provided better performance than models based on a single modality. However, the improvements were not always substantial or statistically significant (Martí-Juan et al., 2020). In the classification of HC vs. AD dementia, accuracies ranged from 90% to 100% (sensitivity 84-100%, specificity 83-100%, AUC 0.96-0.99). Studies differentiating HC from MCI reported accuracies between 74% and 100% (sensitivity 54-100%, specificity 59-100%, AUC 0.77-0.95). In studies differentiating stable MCI cases from progressive MCI cases, accuracies ranged from 57% to 91% (sensitivity 48-96%, 43-95%, AUC 0.69-0.98). In addition, incorporation of longitudinal data improved the performance of the models compared to the models developed using only cross-sectional data (Martí-Juan et al., 2020).

2.6.4.2 Conversion from the preclinical phase to MCI or dementia

As importance of the preclinical AD phase has been recognized during the past years, recent machine learning studies have also focused on the earlier phases of the disease. Gómez-Ramírez et al. (Gómez-Ramírez et al., 2020) searched for the most important self-assessed features for predicting conversion from HC to MCI in a large, observational, longitudinal, Spanish cohort. They utilized Random Forests classifier and permutation-based methods for feature selection and concluded that subjective cognitive decline, educational level, working experience, social life, and diet were among the most important features, subjective cognitive decline being the most important.

Yue et al. (Yue et al., 2021) developed a model for predicting which of the subjects with SCD will progress to MCI. They conducted a feature selection from 233 sociodemographic, health, psychological, and MRI features and fed five most important features to a cost-sensitive SVM developed for imbalanced data sets. The selected features were years of education, Montreal Cognitive Assessment (MoCA) score at baseline, stroke history, volume of left amygdala, and white matter at the banks of right superior temporal sulcus. Their classifier achieved accuracy of 69.7% (sensitivity 62.5%, specificity 73.1%, AUC 0.80).

Liu et al. (Liu et al., 2022, 2020) proposed a joint neuroimage synthesis and representation learning framework for predicting conversion from SCD to MCI. Their method consisted of 1) an image synthesis network which imputes missing images with synthetic images and generates multi-modal imaging features and 2) a classification network which fuses multimodal features and performs prediction. They trained the model with complete MRI and PET scans from the ADNI data set and tested it with the Chinese Longitudinal Aging Study data set containing only MRI images. They reported balanced accuracy of 72.1% (sensitivity 75.0%, specificity 69.2%, AUC 0.75).

Partly the same research group also presented a cost-sensitive meta-learning framework for predicting SCD to MCI conversion using brain MRIs (Guan et al., 2021). This meta learning framework first split training data to a meta-train and meta-test set. The meta-test set consisted of a small sample of SCD subjects from ADNI and the meta-train set included other types of patients (either HC/AD or MCI) from ADNI. A CNN model was developed by jointly optimizing meta-train and meta-test losses. This approach is thought to avoid bias towards the training set (HC/AD or MCI) and learn domain-invariant features. They also introduced a cost-sensitive loss into the meta-learning process to improve sensitivity for detecting progressive SCD cases. This framework was finally evaluated with an independent set of 40 progressive SCD cases and 73 stable SCD cases. It achieved accuracy of 59.7% (balanced accuracy 60.5%, sensitivity 66.4, specificity 54.6%, AUC 0.65).

2.6.4.3 Differential diagnosis

Classification tasks in the previous sections focused on binary classifications along the AD continuum. However, the real-life challenge in memory clinics is to differentiate different forms of dementia from each other. Differential diagnosis in the memory clinics is challenged by 1) overlapping clinical features of different dementias; 2) co-occurrence of different dementias; and 3) less established biomarkers for

dementias other than AD (Bibl et al., 2012; Karantzoulis & Galvin, 2014; Koenig et al., 2018).

Koikkalainen et al. (Koikkalainen et al., 2016) and Tolonen et al. (Tolonen et al., 2018) extended the two-class DSI classifier for a multiclass problem for differentiating subjects with SCD, dementia due to AD, VaD, FTD, and DLB in Amsterdam Dementia Cohort. Koikkalainen et al. (Koikkalainen et al., 2016) used only visually rated and automatically quantified MRI features. The automatic quantification provided better performance than visual ratings (balanced accuracy 69% vs. 52%). With automatic quantification, detection of VaD was the most accurate (sensitivity 96%), followed by SCD (sensitivity 82%), dementia due to AD (sensitivity 74%), and FTD (sensitivity 62%). Detection of DLB was the least accurate (sensitivity 32%). Tolonen et al. (Tolonen et al., 2018) expanded this work by including also data from neuropsychological tests and CSF ($A\beta_{42}$, total tau, p-tau). Addition of two other data modalities increased balanced accuracy to 82%. Similarly, VaD was the most accurately detected (sensitivity 92%), followed by SCD (sensitivity 89%), dementia due to AD (sensitivity 80%), FTD (sensitivity 76%), and DLB (sensitivity 75%). Automatic MRI features produced the best performing single-modality model (balanced accuracy 66%).

Partly the same authors as in the two previous studies developed new atrophy grading and VaD grading MRI features; integrated a sparsity-based multiclass feature selection step; and applied the RUSBoost algorithm to handle the problem of the imbalanced class distribution (Tong et al., 2017). Using the same Amsterdam Dementia Cohort and features from MRI, CSF ($A\beta_{42}$, total tau, p-tau), and age, their method provided the balanced accuracy of 69%. Contrary to the previous studies, SCD was the most accurately detected (sensitivity 82%%), followed by dementia due to AD (sensitivity 81%), VaD (sensitivity 74%%), and FTD (sensitivity 71%). Also in this study, detection of DLB was the least accurate (sensitivity 38%).

Asanomi et al. (Asanomi et al., 2021) utilized penalized regression models (elastic net) with microRNA expression data to differentiate AD, VaD, DLB, normal pressure hydrocephalus, and HC. Their method obtained the balanced accuracy of 37%. HCs were the most accurately detected (sensitivity 79%) and different dementias were considerably less accurately detected (sensitivities <36%).

Bron et al. (Bron et al., 2017) incorporated features from structural MRI, arterial spin labelling, and diffusion tensor imaging to classify HC, AD, and FTD. They trained multiple two-class SVM classifiers (HC vs. AD, HC vs. FTD, AD vs. FTD) and combined them for multiclass classification by multiplying the posterior probabilities. Their method obtained comparable accuracies for AD vs. FTD and HC vs.

AD vs. FTD classifications, the best accuracies being around 75% (estimated from a figure). Several other studies have differentiated only two dementia types from each other or conducted multiple two-class classifications even though they had multiclass data available (Canu et al., 2017; C. Davatzikos et al., 2008; Diehl et al., 2005; Jiménez-Huete et al., 2014; Muñoz-Ruiz et al., 2016; Perani et al., 2016).

2.6.4.4 Prediction of continuous outcomes

As shown above, a large number of studies have focused on predicting categorical outcomes, e.g., differentiating different diagnostic classes or predicting conversion from one disease stage to another. As progression of AD is a continuous process, an alternative approach is to predict continuous outcomes, e.g., future scores or changes in the biomarker values. Studies using continuous clinical scores as outcomes predicted scores or changes in MMSE, ADAS-cog, Clinical Dementia Rating Sum of Boxes (CDR-SB), Clinical Dementia Rating Global (CDR-GLOB), and Rey Auditory Verbal Learning Test (RAVLT) (Bucholc et al., 2019; Huang et al., 2016; Weiner et al., 2017; D. Zhang et al., 2012a, 2012b; Zhou et al., 2013; Zhu et al., 2016). These studies also found that inclusion of multimodal or longitudinal data improved the performance of the models. Especially incorporation of cognitive and functional assessments led to better performance than inclusion of other modalities alone or together (Bucholc et al., 2019; Zhu et al., 2016). Populations in these studies contained only HC, MCI, or AD dementia cases. Li et al. (Li et al., 2022) expanded the population by including also SCD cases in addition to HC and MCI cases. They utilized MRI features from whole brain, hippocampus, and amygdala to predict MMSE and MoCA scores and changes in them. Their method consisted of sparse coding for feature selection and proximity-based Random Forests for regression.

2.6.4.5 Impact of deep learning

As can be seen from the examples in the chapters above, deep learning has been applied to AD in recent years. Contrary to the traditional machine learning methods, deep learning methods do not require a separate feature extraction step before the classification, instead they can process and classify raw images or signals directly. However, deep learning methods require plenty of training data. In AD, the deep learning methods have improved MRI as a single data modality (Weiner et al., 2017). The deep learning models based only on MRI have achieved accuracies between

92% and 97% in HC vs. AD classification and over 80% in MCI to AD conversion. Nevertheless, multimodal based deep learning approaches have shown the best performance (Jo et al., 2019; Weiner et al., 2017). In addition, hybrid models, which use deep learning for feature extraction and traditional machine learning methods for classification, have provided promising results with limited data sets (Jo et al., 2019).

2.6.5 Challenges and limitations

This section describes challenges, limitations, and issues to consider in research on machine learning in AD. First, there are always missing values in healthcare data sets. Especially in longitudinal data, not all tests and measurements are available for all subjects at all time points. There are several ways to overcome the problem of missing values: restricting analyses to the cases or variables that have all data available (complete case analysis); using existing values in variables being studied, instead of deleting all cases with missing values in other variables (pairwise analysis); imputing missing values; and utilizing methods that tolerate missing data (Acock, 2005; Kang, 2013).

Second, definite diagnosis of AD is based on histopathological postmortem samples. Thus, the diagnoses available are not always correct, and they may change as the time goes by and the patient state develops. That is why long follow-up times are needed to obtain as reliable diagnoses as possible. In addition, the diagnostic guidelines have evolved and changed over the years, which may affect results if data sets from different time points are used in the development and evaluation of the methods. This also hampers interpretation and comparison of the results from the past and present. Furthermore, different centres and countries may have different practices for the diagnostic process, e.g., the diagnosis is done by one clinician or a group of professionals with different backgrounds. There is also variation between the clinicians in interpretation of the patient data and confidence in giving the diagnosis, which may lead to differing diagnoses. Especially, there might be more variation in the diagnosis of not so clear cases.

Third, most of the machine learning studies in AD utilized cross-validation in performance evaluation, and a few studies used independent data sets for final evaluation. In addition, the ADNI data set was the most used data set. This limits generalizability of the developed models and increases the risk of overfitting if the models are complex, many parameters are tuned, and features selected. This also has the risk that the research community is overtraining their models on this one specific

dataset and for which only the 'best suited' results are published (Gencoglu et al., 2019).

Fourth, many machine learning methods, especially deep learning methods, are black boxes, meaning that they do not reveal why a certain result was given or which variables were the most influential ones. In addition, mathematics behind the model may be difficult to understand for people without knowledge on machine learning. This is problematic in the healthcare domain where interpretability, understandability, and transparency of the methods is important. This requirement has led to development of interpretable machine learning methods. The DSI method and the related DSF visualization are examples of such interpretable methods, and they are described in detail in Chapters 4.2.1 and 4.2.2. Qiu et al. (Qiu et al., 2020) presented another interpretable approach using fully convolutional network to create disease probability maps from brain MRI scans. The probability maps visualized AD risk in different brain regions and were used as inputs for multilayer perceptron for final classification of HC and AD cases. Das et al. (Das et al., 2019) proposed a sparse high-order interaction model with rejection option (SHIMR). This approach creates a set of simple and interpretable if-then rules. These rules are then visualized highlighting the variables forming each rule and the importance (weight) of each rule in forming an overall score. In addition, pointers show subject's variable values and overall score on the scale. The rejection option of the SHIMR model declines to give decisions when the diagnosis is not confident enough, i.e., the result is close to the decision boundary. The SHIMR model was tested with plasma proteomics and CSF biomarkers to differentiate HC from AD cases.

3 AIMS OF THE STUDY

The objective of this thesis was to develop and validate data-driven methods for predicting and monitoring progression of Alzheimer's disease at the different phases of the disease spectrum, starting from normal cognition and ending to death, using data from neuropsychological and cognitive tests, MRI, CSF, comorbidities, and APOE. Specific aims of the thesis were:

O1. to develop a model for predicting who of the individuals with subjective cognitive decline are at risk for mild cognitive impairment or dementia (Study I)

O2. to develop methods for monitoring progression of disease over time in a mild cognitive impairment cohort in which some progressed to dementia due to Alzheimer's disease and others did not (Study II)

O3. to predict atrophy of hippocampus in a population consisting of subjects with normal cognition, mild cognitive impairment, and dementia due to Alzheimer's disease (Study III)

O4. to identify which of disease-related determinants are associated with mortality in patients with dementia due to Alzheimer's disease (Study IV)

Figure 4 illustrates how the objectives link to the disease spectrum and relate to each other.

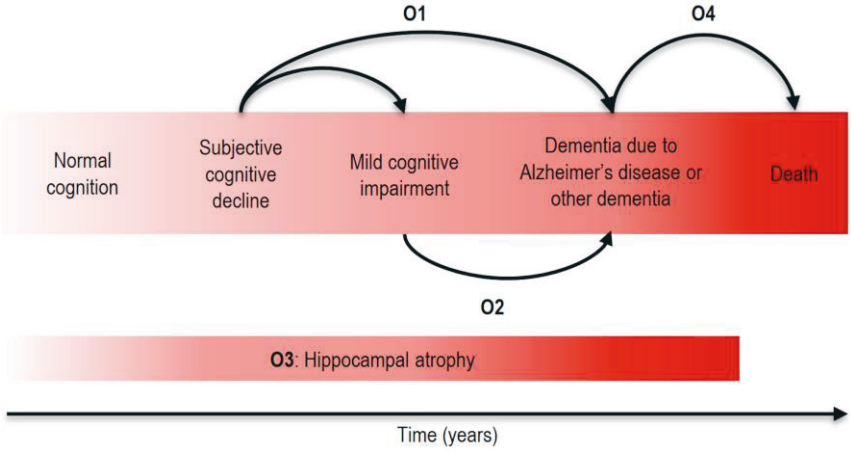


Figure 4. Summary of the objectives

4 MATERIALS AND METHODS

This chapter describes the materials and methods used in the four original publications. First, the study populations and available variables are presented. Then, modelling and performance evaluation methods are described.

4.1 Study populations and variables

Table 2 summarizes the characteristics of the study populations in Studies I-IV. It also shows the data modalities used in the analyses. The populations and used variables are described in more detail in the following sections.

Table 2. Summary of the study populations

	Study I			**Study II**	**Study III**		**Study IV**
Name	ADC	DCN	BAR	ADNI 1	ADNI 1	AIBL	ADC
Location	Nether-lands	Germany	Spain	USA Canada	USA Canada	Australia	Netherlands
Subjects	354	269	51	273	530	176	616
Diagnosis at the baseline	SCD	SCD	SCD	MCI	HC MCI AD	HC MCI AD	AD
Demographics	x	x	x	x	x	x	x
Cognitive and neuro-psychological tests	x	x	x	x	x		x
CSF							
Aβ	x	x	x	x	x		x
Total tau	x	x	x	x	x		x
P-tau	x	x	x		x		x
MRI							
Automatic features	x	x	x	x	x	x	
Visual rating							x
APOE	x	x	x	x	x		x
Comorbidities							x

4.1.1 Prediction of progression of SCD (Study I)

Study I was based on three memory clinic cohorts: Amsterdam Dementia Cohort (ADC) collected in VU Medical Center in Amsterdam (van der Flier et al., 2014), German Dementia Competence Network (DCN) collected in nine memory clinics in Germany (Kornhuber et al., 2009; Wolfsgruber et al., 2017) and a cohort collected in Barcelona (BAR) (Valech et al., 2015). From the three cohorts, Study I included in total 674 subjects with SCD, having neuropsychology at baseline available and a minimum follow-up of one year. The subjects were divided into two groups: stable SCD (SSCD), whose diagnosis stayed as SCD during the follow-up, and progressive SCD (PSCD), who progressed to MCI or dementia during the follow-up. In total, 22% of the population had progressive SCD. Table 3 presents characteristics of the population.

Table 3. Characteristics of the population in Study I

	ADC		**DCN**		**BAR**	
	SSCD	PSCD	SSCD	PSCD	SSCD	PSCD
Subjects, n (%)	291 (82)	63 (18)	186 (69)	83 (31)	46 (90)	5 (10)
Female, n (%)	138 (47)	26 (41)	71 (38)	34 (41)	34 (74)	4 (80)
Age, years	61.2 ± 9.6	69.0 ± 7.1	64.5 ± 7.8	68.0 ± 8.4	64.9 ± 6.4	70.2 ± 8.3
Education, years	13.3 ± 4.3	14.0 ± 4.4	12.5 ± 2.8	13.3 ± 3.3	10.8 ± 4.2	11.6 ± 4.3
Follow-up, years	3.4 ± 2.2	3.8 ± 3.2	2.3 ± 0.9	1.6 ± 0.7	3.7 ± 1.8	2.8 ± 1.8
MMSE at baseline	28.4 ± 1.7	28.0 ± 1.5	28.2 ± 1.6	27.6 ± 1.8	28.3 ± 1.5	26.8 ± 1.9
MCI/AD dementia/other dementia, n		42/15/6		53/21/9		2/2/1

Values are expressed as mean ± standard deviation unless otherwise stated.

The data used in Study I included neuropsychology, CSF, APOE, and MRI. Neuropsychological functioning was evaluated using standardized tests overlapping between the three centres: MMSE, Trail Making Test A (TMT-A) and Test B (TMT-B), and Category Fluency (animals). Episodic memory was examined with different tests in the different centres, thus, the tests resembling each other the most were selected: RAVLT in ADC, Consortium to Establish a Registry for Alzheimer's Disease (CERAD) in DCN, and Free and Cued Selective Reminding Test (FCSRT) in BAR. CSF included Aβ, total tau, and p-tau. APOE status was recorded as ε4 carrier or non-carrier. Four features were automatically extracted from MRI scans using an image quantification tool: hippocampal volume, computed medial temporal lobe atrophy score (cMTA), computed global cortical atrophy score (cGCA), and region of interest (ROI) based grading. cMTA and cGCA are automatically computed versions

of medial temporal lobe and global cortical atrophy which are commonly rated visually by radiologists. ROI-based grading measures similarity of a ROI in an image from a subject being studied to corresponding ROIs in images in a reference database. In this case, the ROI was centred on the hippocampus.

4.1.2 Monitoring progression of MCI (Study II)

Data utilized in Study II were obtained from the publicly available ADNI 1 study (see details in Section 2.6.2). Study II included 289 subjects with MCI having at least 24 months of follow-up data available. The subjects were classified into two groups: a stable MCI (SMCI), whose diagnosis stayed as MCI during the follow-up, and a progressive MCI (PMCI), whose diagnosis changed from MCI to AD during the follow-up. Table 4 presents characteristics of the two subject groups.

Table 4. Characteristics of the population in Study II

	SMCI	PMCI
Subjects, n (%)	149 (51.6)	140 (48.4)
Gender, n (%)		
Male	51 (34.2)	55 (39.3)
Female	98 (65.8)	85 (60.7)
Age, years	75.1 ± 7.4	75.4 ± 6.7
Education, years	15.9 ± 3.0	15.6 ± 3.0

Values are expressed as mean ± standard deviation unless otherwise stated.

The data utilized in Study II comprised neuropsychology, CSF, APOE, and MRI. Neuropsychology included MMSE, ADAS-cog, clock draw and copy, Logical Memory I and II, RAVLT, Digit Span Forward and Backword, Category Fluency, TMT-A, TMT-B, Digit Symbol Substitution Test, Boston Naming Test, and American National Adult Reading Test. CSF consisted of Aβ and total tau, and APOE was recorded as a type of alleles (ε2, ε3, or ε4). MRI included 14 features measuring volumes of different brain regions. The features were provided to the ADNI site by Anders Dale Lab (University of California). A detailed list of the MRI features can be found in Supplementary Material of Publication II.

4.1.3 Prediction of hippocampal atrophy (Study III)

Data for Study III were obtained from the ADNI 1 and AIBL databases. ADNI 1 was already described in Section 2.6.2, thus, only AIBL is shortly described here. AIBL aimed at discovering biomarkers, cognitive characteristics, and health and lifestyle factors affecting development of symptomatic AD (Ellis et al., 2009; Fowler et al., 2021). It started in 2006 and recruited 1 112 subjects (HC, MCI, dementia due to AD) of at least 60 years (Ellis et al., 2009; Fowler et al., 2021). The subjects have been followed for 126 months at 18 months intervals (Fowler et al., 2021). The study is conducted in Perth in Western Australia and Melbourne in Victoria (Ellis et al., 2009; Fowler et al., 2021). The data from the AIBL participants having ADNI-compliant PET and MRI, are available through the ADNI website. This population consists of about 25% of the full AIBL cohort.

Study III consisted of ADNI 1 and AIBL subjects having both baseline and follow-up MRI available (24-month MRI in ADNI 1, 18-month MRI in AIBL). Four sets of analyses were conducted for development and validation of the models. Analysis 1 included the full set of variables described later in this chapter. As only a subset of subjects had CSF biomarkers available, the population in Analysis 1 was restricted to those having the CSF biomarkers. Analysis 2-4 included only MRI features. In Analysis 2, the models were trained using all subjects with both baseline and follow-up MRI, but for the comparison purposes, the results were only shown for the same subjects as in Analysis 1. Analysis 3 included all subjects with baseline and follow-up MRI. In Analysis 4, the models were trained using ADNI 1 and tested using AIBL. All data modalities were not available in AIBL, thus the models including the full set of variables could not be evaluated on the independent test data. Table 5 presents characteristics of the populations in these different analyses.

Table 5. Characteristics of the population in Study III

	ADNI 1		AIBL
	Analysis 1 & 2	**Analysis 3**	**Analysis 4**
Subjects, n	281	530	176
Age, years	75.1 ± 6.7	75.4 ± 6.5	71.9 ± 7.2
Gender, n (%)			
Male	167 (59.4)	309 (58.3)	88 (50.0)
Female	114 (40.6)	221 (41.7)	88 (50.0)
MMSE at baseline	26.8 ± 2.6	27.0 ± 2.6	27.8 ± 2.9
Diagnosis, n (%)			
HC	75 (26.7)	151 (28.5)	119 (67.6)
SMCI	61 (21.7)	118 (22.3)	12 (6.8)
PMCI	64 (22.8)	118 (22.3)	6 (3.4)
AD	63 (22.4)	109 (20.6)	21 (11.9)
Unknown	18 (6.4)	34 (6.4)	18 (10.2)

Values are expressed as mean ± standard deviation unless otherwise stated.

Study III included similar data to Study II: neuropsychology, CSF, APOE, and MRI. Neuropsychology comprised MMSE, ADAS-cog, Clinical Dementia Rating (CDR), clock draw and copy, RAVLT, Digit Span Forward and Backward, Category Fluency, TMT-A, TMT-B, Digit Symbol Substitution Test, Boston Naming Test, and American National Adult Reading Test. CSF consisted of Aβ, total tau, and p-tau, and APOE was recorded as a type of alleles (ε2, ε3, or ε4). Volumes of different brain regions were extracted from the MRI using volumetric segmentation (Lötjönen et al., 2010), tensor-based morphometry (TBM) (Koikkalainen et al., 2011), and voxel-based morphometry (VBM) (Ashburner & Friston, 2000). Rate of hippocampal atrophy was computed using a method called extended boundary shift integral (Lötjönen et al., 2014). ADNI 1 had follow-up MRI available at months 12 and 24 and AIBL at month 18. Atrophy rate in AIBL was multiplied by the factor of 2/3 to make it consistent with the annual atrophy rate at month 24 in ADNI 1. All MRI features are listed in Supplementary Material of Publication III.

4.1.4 Factors associated with mortality in AD (Study IV)

As Study I, Study IV also utilized the ADC (van der Flier et al., 2014), from which 616 subjects with dementia due to AD were selected for the analyses. Inclusion cri-

teria were as follows: a baseline visit between years 2000 and 2014, neuropsychological test battery available at the baseline, MMSE score ≥16, and a minimum follow-up of 2 years. Information on deaths were obtained from the Dutch Municipal Register. Table 6 presents characteristics of the selected subjects.

Table 6. Characteristics of the population in Study IV

	Alive	Died
Subjects, n (%)	403 (65)	213 (35)
Female, n (%)	218 (54)	91 (43)
Age, years	66 ± 7	69 ± 9
Education, years	11 ± 3	11 ± 3
MMSE at baseline	22 ± 3	22 ± 3
Follow-up, years	5.3 ± 1.8	4.3 ± 2.1

Values are expressed as mean ± standard deviation unless otherwise stated.

Study IV contained variables describing demographics, medical history, neuropsychology, CSF, and MRI. Demographics included gender, age, years of education, years of complaints, APOE ε4 status (carrier, non-carrier), and activities of daily living. Medical history was recorded and defined using number of medications, smoking status, and presence (yes/no) of hypertension, hypercholesterolemia, diabetes mellitus, and cardiovascular disease. Neuropsychology was assessed utilizing MMSE, Digit Span Forward and Backward, Visual Association Test (VAT), RAVLT, TMT-A, TMT-B, and Category Fluency (animals). CSF consisted of Aβ, total tau, and p-tau. MRI scans were visually rated by trained assessors and afterwards evaluated in a consensus meeting with an experienced neuroradiologist. Visual ratings included medial temporal lobe atrophy (MTA, range 0-4), global cortical atrophy (GCA, range 0-3), parietal atrophy (PA, range 0-3, average over the left and right), white matter hyperintensities (WMH, range 0-4), presence of lacunes, presence of infarcts, and number of microbleeds (categorized to 0, 1-2, ≥3).

4.2 Modelling methods

4.2.1 Disease State Index

Study I and Study II utilized the DSI method for predicting and monitoring disease progression. DSI is a supervised machine learning method, combining information

in several features to quantify the state of a disease in a patient. A DSI value for a subject is computed by comparing the subject's measurement values to data measured from known healthy (control) and diseased cases. DSI provides as an output a numeric value between [0, 1], higher values indicating greater similarity of the subject to the diseased cases on the basis of the measured data. Thus, higher values denote more advanced disease state.

The DSI computations are based on the differences in the probability density functions (PDFs) of measurement values between the healthy and diseased populations (e.g., healthy and AD). Two distinct values are defined using the PDFs: fitness and relevance. The fitness value is in the range [0, 1] and it evaluates similarity of the subject's measurement value to these two groups. It is defined as a likelihood of the measured value belonging to the diseased group:

$$fitness(x) = \frac{FN(x)}{FN(x) + FP(x)}, \tag{1}$$

where FN(x) is the false negative error rate and FP(x) the false positive error rate in the training data when using subject's measurement value x as a classification threshold.

Some features are better in the discrimination of the two groups than other features. In the DSI, this is taken into account with the relevance value. The relevance measures the feature's ability to separate the two groups from each other. Like the fitness values, the relevance values are also in the range [0, 1], with increasing values indicating a better discrimination ability. The relevance is zero if the PDFs overlap entirely (poor discrimination) and one if there is no overlap (perfect discrimination). The relevance is computed as

$$relevance = sensitivity + specificity - 1. \tag{2}$$

A composite DSI value, combining several features, is computed as a weighted average where the fitness values are weighted according to their relevance:

$$DSI = \frac{\sum relevance \times fitness}{\sum relevance}. \tag{3}$$

These computations (fitness, relevance, averaging) can be performed several times recursively to obtain a hierarchy of the DSI values to describe how different features contribute to the overall result. This hierarchy can be visualised with DSF, which is described in the next section. More details on the DSI computations can be found in a publication by Mattila and co-workers (Mattila et al., 2011).

Study I utilized the DSI method for predicting which of the subjects with SCD progressed to MCI or dementia. The subjects who did not progress were regarded

as a control population and the subjects who progressed were regarded as a diseased population in the DSI computations.

In Study II, the DSI values were calculated at different time points for the population consisting of subjects with SMCI or PMCI. As the patients entered the study at the different phases of the disease, the time stamps of their visits were first synchronized. For the PMCI cases, the moment of receiving the AD diagnosis was set as a time point zero (Z). Whereas for the SMCI cases, the last available time point until month 36 was set as the Z. The preceding time points were labelled as Z-6, Z-12, etc. Longitudinal behaviour of the DSI values was studied by performing linear least mean square (LMS) regression of the DSI values over the synchronized time points for each subject separately. Trend parameters of the regression (slope, intercept) were compared between the SMCI and PMCI groups to study whether the DSI values reflected the progression of the disease over time.

4.2.2 Disease State Fingerprint

DSF provides a visual representation of patient data and the hierarchy of the DSI values, summarizing how different features contribute to the total DSI value. An example DSF is shown in Figure 5. DSF is composed of a tree structure with nodes in different sizes and colours. The size of the node is determined by the relevance value: the bigger the node, the more important the feature. Sibling nodes of the tree are sorted in decreasing order by their relevance values. The colour of the node reflects the DSI value: shades of red refer to higher DSI values (similarity to diseased cases) and shades of blue refer to lower DSI values (similarity to control cases). Leaf nodes show the individual features. Numbers in the parentheses show the DSI value. Grouping of the features into different categories (e.g., neuropsychological or MRI) is defined by the user. In Study I, the DSF methodology was utilized for visualizing data from a single time point (baseline). In Study II, the DSF was extended for visualizing data from multiple time points to allow follow-up of the disease progression.

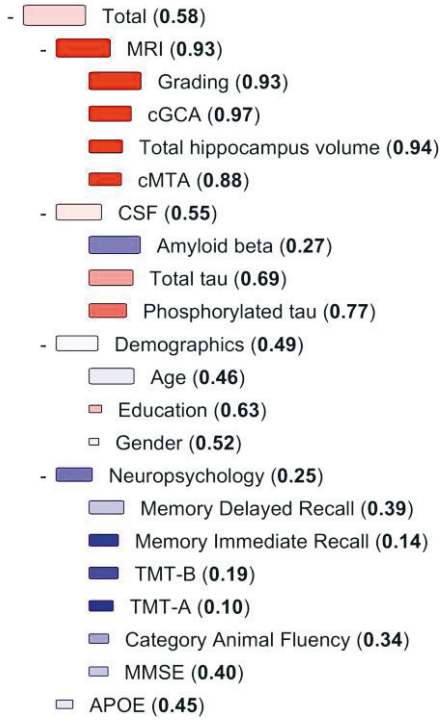


Figure 5. An example of a Disease State Fingerprint visualization.

4.2.3 Naïve Bayes

Study I compared performance of the DSI method to the Naïve Bayes (NB) classifier in prediction of progression of SCD to MCI or dementia. The NB classifier is a traditional classifier, which is based on Bayes' rule defined as

$$p(\omega_k|x) = \frac{p(x|\omega_k)P(\omega_k)}{p(x)}, \tag{4}$$

where ω_k is the kth class and x is the feature vector describing the object to be classified (Duda et al., 2001; Tohka & van Gils, 2021; Webb & Copsey, 2011). *The posterior probability* $p(\omega_k|x)$ is the probability of the class ω_k to be the correct class for the object given the feature vector x describing the object. *The class conditional PDF* $p(x|\omega_k)$ defines the PDF of the feature vector x given the class ω_k. *The prior probability* $P(\omega_k)$ of the class ω_k is percentage of the of all objects belonging to the class ω_k. The function p(x) is defined as

$$p(x) = \sum_{j=1}^{C} p(x|\omega_k)P(\omega_k). \tag{5}$$

The function p(x) is same for all classes and can be seen as a scaling factor which guarantees that the posterior probabilities sum to one. As the Bayes classifier selects the class that has the greatest posterior probability, p(x) can be dropped from Eq. 4. Thus, the Bayes classifier is defined by the class conditional PDFs $p(x|\omega_k)$ and prior probabilities $P(\omega_k)$. The class conditional PDFs can be estimated using parametric methods, which assume PDFs to have a certain parametric form (e.g., Gaussian distribution), or non-parametric methods, which do not make any assumptions about the form of the PDFs (e.g., Parzen window). The NB classifier is based on a naïve assumption that features are statistically independent given the class to which the feature vector belongs. In Study I, the class conditional PDFs of continuous variables were estimated with Gaussian distribution:

$$N(x;\ \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\ \mu)^2}{2\sigma^2}\right), \tag{6}$$

where x is a value of the feature, μ is mean of the feature and σ2 is variance of the feature (Webb & Copsey, 2011). When applying Bayes' rule to categorical data, the PDFs are replaced by probabilities. In Study I, the class conditional probabilities of categorical variables were estimated using additive smoothing:

$$P(x = L|\omega_k) = \frac{1 + n_{Lk}}{n_l + n_k}, \tag{7}$$

where L is a category of the feature, $n_{L,k}$ is the number of occurrences of category L in class ω_k, n_l is the total number of categories, n_k is the total number cases in class ω_k (Manning et al., 2008; MathWorks, 2021). Additive smoothing solves the zero frequency problem which arises when a certain category is present in the test data but not in the training data (Manning et al., 2008). Finally, the NB classifier chooses the class that maximises the product of the prior probability and class conditional probability densities (or probabilities) of the individual features:

$$\hat{y} = \max_{k=1,\ldots,C} P(\omega_k) \prod_{m=1}^{D} p(x_m|\omega_k), \tag{8}$$

where D is the total number of features. For continuous features, $p(x_m|\omega_k)$ is the class conditional Gaussian PDF. For categorical features, $p(x_m|\omega_k)$ is the class conditional probability.

4.2.4 Random Forests

Study I also compared performance of the DSI method to Random Forests (RF) classifier. The RF classifier is an ensemble classifier which trains several decision trees and the final classification is based on majority voting of the trees. Each individual tree is grown by 1) drawing a bootstrap sample of size n by sampling with replacement from the training data of size n; 2) selecting m features at random from the bootstrap sample (m<<number of input features); 3) picking the best split on the m features for splitting the node of the tree; 4) splitting the node into two daughter nodes on the basis of the best split; 5) growing the tree to the largest extent possible without any pruning (Breiman, 2001; Hastie et al., 2009). The RF classifiers do not require performance evaluation with a separate test set or cross-validation as the bootstrap method utilises about two-thirds of the cases for growing of the tree and excludes the remaining one-third, called the out-of-bag sample, which is used for calculation of the error estimate (Breiman, 2001; Hastie et al., 2009).

4.2.5 Regularized regression

Study III predicted change in hippocampal volume during 24 months using baseline data. Regression was chosen as a modelling method because hippocampal volume is a continuous variable. When the number of features is large, the normal LMS regression is not practical because it includes all features into the model. Instead, regularized regression was chosen since it includes penalty term, constraining the size of the estimated coefficients. Thus, regularized regression sets some of the coefficients to zero, which leads in practice to features selection. Study III utilized regularized regression with the elastic net regularization defined as

$$\min_{\beta_0,\beta}\left(\frac{1}{2N}\sum_{i=1}^{N}(y_i-\beta_0-x_i^T\beta)^2+\lambda P_\alpha(\beta)\right), \tag{9}$$

where

$$P_\alpha(\beta)=\frac{(1-\alpha)}{2}\|\beta\|_2^2+\alpha\|\beta\|_1=\sum_{j=1}^{D}\frac{(1-\alpha)}{2}\beta_j^2+\alpha|\beta_j|, \tag{10}$$

where β_0 is intercept, β is a vector of regression coefficients, N is number of subjects, y_i is the response (observed outcome) of the subject i, x_i is the vector including predictor data for the subject i, λ is a positive regularization parameter (higher λ, less

non-zero coefficients), α defines the weight of L1 and L2 norms, D is the number of predictors in the model (Friedman et al., 2010; Zou & Hastie, 2005). The parameter α can have values between zero and one. When α is close to zero, the model approaches the ridge regression, emphasizing the L2 norm. When α is one, the model represents the lasso regression, involving the L1 norm. For other values of α, the penalty term interpolates between the L1 and squared L2 norms of β. In Study III, α was set to 0.5 and 1.0.

4.2.6 Least absolute deviation regression

A drawback of the LMS regression is its sensitivity to outlying observations. A more robust method towards the outlying observations is the least absolute deviation (LAD) regression. Instead of minimizing the sum of squared residuals as in the LMS regression, the LAD regression minimizes the sum of absolute values of the residuals:

$$\min_{\beta_0,\beta}\left(\sum_{i=1}^{N}|y_i-\beta_0-x_i^T\beta|\right), \tag{11}$$

where β_0 is intercept, β is a vector of regression coefficients, N is number of subjects, y_i is the response (observed outcome) of the subject i, x_i is a vector including predictor data for the subject i (Dasgupta & Mishra, 2004; Dielman, 2005). In addition to the regularized LMS regression, Study III utilized the LAD regression. As LAD does not include regularization or feature selection, the regularized LMS regression was performed first and then features with non-zero coefficients were fed into the LAD regression.

4.2.7 Cox proportional hazards model

Study IV used Cox proportional hazards (CPH) model to evaluate associations between different predictors and mortality. As opposite to LMS regression, the CPH model takes both time to an event (e.g. death) and predictors into account. The CPH model is written as

$$h(t)=h_0(t)\times\exp\left(\sum_{i=1}^{D}b_ix_i\right), \tag{12}$$

where t is time to an event, h(t) is the hazard function at time t, $h_0(t)$ is the baseline hazard, x_i is a predictor i, b_i is a coefficient of the predictor i, and D is the number of predictors (Bradburn et al., 2003). The hazard function measures probability of a subject to experience the event at time t. The baseline hazard is the value of the hazard when all predictor values are equal to zero. It is estimated non-parametrically, thus, survival times are not assumed to follow any particular statistical distribution. The terms $\exp(b_i)$ are called hazard ratios (HR) and they measure the impact of the predictor. HR above one indicates positive association between the predictor and event: when the value of the predictor increases, the hazard of the event increases also. HR below one indicates negative association: when the value of the predictor increases, the hazard of the event decreases. HR of one indicates no association.

In addition to investigating associations of individual predictors with mortality, Study IV also searched for an optimal combination of predictors. For this, a multivariable model was constructed using forward selection. The forward selection starts with an empty model and adds in predictors with the lowest p-value one by one until $p<0.10$. The predictors were added only if the overall model improved.

4.3 Performance evaluation

Table 7 summarizes all modelling and performance evaluation methods used in Studies I-IV. Study I focused on predicting who of the subjects with SCD will progress to MCI or dementia. The developed models were trained and internally evaluated using the ADC cohort and 10 iterations of stratified 3-fold CV. External evaluation was performed by training the models with the whole ADC cohort and testing them with the pooled DCN and BAR cohorts. To understand reasons for decreased performance in the external validation, the evaluation was repeated by performing CV in the pooled DCN and BAR cohorts and using the ADC cohort as an independent test set. Study I compared performance of the DSI classifiers to the NB and RF classifiers. As classifiers like NB and RF may perform poorly on imbalanced class distributions, 1) the original data set was used as such, and 2) the number of progressors was increased to match the number of stable cases by duplicating randomly selected progressors and adding them to data set. Some of the patients in ADC were scanned with 1.0 T MRI devices and others with 1.5 T or 3.0 T devices. To reduce heterogeneity caused by the different field strengths, the MRI features of the patients scanned with the 1.0 T devices were excluded in the training phase. Testing was performed using MRI features from all field strengths and only >1.0 T.

Study II investigated longitudinal behaviour of the DSI values in subjects with stable or progressive MCI. The DSI values were calculated using stratified 10-fold CV. The training set contained data from the baseline for SMCI cases and data at the time of conversion for the PMCI cases. The test sets included data from all available visits of the remaining cases. This selection of the training data sets the dynamic range of the DSI method between the SMCI cases at the baseline and early AD. To have complete data sets at each visit, missing values were imputed with the patient's previous available value. This may lead to conservative disease progression estimates for some patients.

Study III comprised four sets of analyses described in Section 4.1.3. Prediction performance of the models in Analyses 1-3 was evaluated using the ADNI 1 cohort and the nested CV. In the nested CV, the outer CV was the 10-fold CV, stratified according to four diagnostic classes (HC, SMCI, PMCI, AD). The inner CV defined the optimal value for the regularization parameter λ in the stratified 5-fold CV. In Analysis 4, generalizability of the models was evaluated by training the models with the ADNI 1 data and testing them with the independent test data from the AIBL study. Performance was measured with RMSE and Spearman correlation coefficient. In addition, the continuous outcome was dichotomized to evaluate how well the models predict which of the subjects have fast or slow rate of hippocampal atrophy. The threshold for dichotomization was set to the middle point between the average atrophy rates for HC and AD. The number of cases in each diagnostic class was almost equal in ADNI 1, whereas AIBL included nearly six times more HCs than subjects with AD. The imbalance was taken into account by randomly selecting 20 HCs with all subjects in other diagnostic classes and then calculating the performance metrics. Number of 20 was chosen because it was close to the number of AD cases in AIBL. This process was iterated 20 times. Missing values in the predictor variables were imputed with the medians, which should only have a minor effect on the results because more than 99% of the variables had at most three missing values in the different analysis sets.

Study IV utilized the CPH models to evaluate associations of baseline variables with mortality. The strength of the associations was measured with HRs and accompanying 95% confidence intervals (CI). Thus, Study IV did not include cross-validation or other performance evaluation methods. Missing values were imputed using multiple imputation, which estimated the missing values on the basis of other available baseline variables in 15 imputation cycles.

Table 7. Summary of the modelling and performance evaluation methods

	Study I	**Study II**	**Study III**	**Study IV**
Model	DSI/DSF RF NB	Linear LMS regression of the DSI values over time	Model 1: regularized LMS regression Model 2: feature selection with regularized LMS regression and modelling with LAD regression	Cox proportional hazards model Strengths of associations evaluated with HRs and 95% CI
Validation	10 iterations of stratified 3-fold CV in ADC Training with ADC and testing with the pooled DCN and BAR 10 iterations of stratified 3-fold CV in the pooled DCN and BAR Training with the pooled DCN and BAR and testing with ADC	Stratified 10-fold CV	Nested CV in ADNI 1: *Outer CV*: Stratified 10-fold CV *Inner CV*: stratified 5-fold CV for defining the parameter λ Training with ADNI 1 and testing with AIBL	NA
Performance metrics	AUC BACC Sensitivity Specificity NPV PPV	AUC Accuracy Sensitivity Specificity	Continuous outcome: RMSE Spearman correlation coefficient Dichotomized outcome: Accuracy Sensitivity Specificity NPV PPV	NA

5 RESULTS

5.1 Prediction of progression of SCD (Study I)

Table 8 shows performance of the classifiers to predict conversion from SCD to MCI or dementia using CV in the ADC cohort. When using only one data modality, the CSF-based DSI classifier had the highest AUC and BACC, followed by demographics- and MRI-based classifiers. When all data modalities were used together, the performance of the DSI classifier improved (AUC 0.81, BACC 74%). The results of the classifiers including MRI features were separately shown for all field strengths and field strengths of >1T. When using only MRI features, the >1T results had somewhat higher AUC and BACC and considerably higher specificity, whereas all field strengths had higher sensitivity. When all data modalities were used together, other features compensated the differences caused by the different field strengths. AUC and BACC between the different field strengths were now similar and specificity for all field strengths improved. Sensitivity for >1T improved also, but it was still somewhat lower than for all field strengths. All single- and multimodality DSI classifiers had high NPV (>88%) and low PPV (<38%). These analyses were repeated for predicting conversion to MCI or dementia due to AD (excluding other dementias, n=16) and results were comparable (results not shown here).

Table 8 also presents performance of the DSI classifier for a subgroup of patients having extreme DSI values calculated using all features. The subgroup consisting of patients with DSI < 0.3 or DSI > 0.7 included almost half of the population. For this subgroup, all performance measures improved when compared to the whole population. NPV was especially high (97%), whereas PPV was only modest (51%).

Finally, Table 8 shows performance of the NB and RF classifiers using all data modalities together and all field strengths. When the original data set was used as such, both NB and RF performed somewhat worse than DSI. When the number of progressors was increased by randomly duplicating progressive cases, NB provided comparable results to DSI and RF still performed somewhat worse. The differences between NPV and PPV provided by NB and RF were smaller than provided by DSI.

Table 9 shows performance of the classifiers when evaluated using the independent test set consisting of the pooled DCN and BAR cohorts. The best performing

data modality was MRI (AUC 0.77, BACC 67%). Performance of CSF decreased, but only 40% of the test population had CSF available. This is considerably smaller proportion than in the ADC cohort, in which 64% of the population had CSF. Performance obtained using all features decreased in the independent test set compared to the cross-validated results on the ADC cohort (AUC 0.72, BACC 65%, NPV 84%). To understand reasons for the decrease, the performance evaluation process was repeated other way around, i.e., using the pooled data from DCN and BAR as the training set and the ADC data as the test set. The same decrease was seen in the cross-validated results on the pooled BAR and DCN cohorts using all features (AUC 0.73, BACC 68%). When the ADC cohort was used as the independent test set, performance was as good as the cross-validated result on the ADC data (all features and all field strengths: AUC 0.79, BACC 73%; all features and MRI >1T: AUC 0.81, BACC 73%). Same decrease in the performance was also seen in the NB and RF classifiers. The lower performance may be attributable to the several differences between the cohorts: 1) CSF, which was the best performing data modality in the ADC cohort, was available only in the small proportion of subjects in DCN and BAR; 2) different criteria for MCI diagnosis (Petersen criteria in ADC and BAR, Jak-Bondi criteria in DCN); 3) differences in the baseline characteristics (patients in ADC were younger, patients in BAR were more often female and had less education); 4) differences in the follow-up durations and progression rates (longer follow-up time in ADC, more progression in the pooled DCN and BAR); 5) different memory tests were used in the cohorts; 6) APOE genotype was determined using different methods; 7) CSF samples from ADC and BAR were analysed in the same laboratory, whereas DCN used a different laboratory.

Figure 6 presents examples of the DSF visualizations for three patients. Patient A has a total DSI value of 0.20 and almost all boxes in the DSF are blue, indicating that the patient is not likely to progress and can be reassured. This patient remained stable during a 3-year follow-up. Patient B has a total DSI of 0.83, mainly caused by the patient's values on MRI and CSF (shown as red boxes in the DSF). This suggests that risk of progression is increased, and follow-up should be considered. This patient progressed to MCI after three years. Patient C has a total DSI of 0.47 and boxes in the DSF have different shades of red and blue. This indicates inconclusive results and reliable prognosis cannot be made. This patient remained stable during a 4-year follow-up.

Table 8. Classification performance of the classifiers evaluated using cross-validation in the ADC cohort

Classi-fier	Variable	%	SSCD, n	PSCD, n	AUC	BACC	Sensitivity	Specificity	PPV	NPV
DSI	Demographics		291	63	0.74±0.04	66.0±5.0	66.0±11.7	65.9±6.1	29.7±3.8	90.1±2.8
	APOE		267	50	0.60±0.05	59.7±4.9	53.9±8.4	65.5±4.5	22.7±4.4	88.4±2.4
	Neuropsychology		290	62	0.69±0.06	62.7±4.4	61.6±10.8	64.3±5.3	26.9±3.3	88.7±2.4
	CSF		194	33	0.77±0.07	69.9±5.0	66.1±11.3	73.6±6.5	30.3±5.8	92.8±2.6
	MRI (1T, 1.5T, 3T)		277	55	0.68±0.05	61.4±4.3	80.1±10.1	42.8±6.8	21.8±2.5	91.9±3.5
	MRI (>1T)		123	25	0.73±0.09	69.1±7.8	64.9±15.2	73.3±7.6	33.6±9.6	91.3±4.0
	All features (1T, 1.5T, 3T)		291	63	0.80±0.05	74.0±4.2	82.9±8.4	65.1±5.8	34.2±3.8	94.7±2.4
	All features (>1T)		291	63	0.81±0.06	74.1±5.8	75.7±11.2	72.6±4.8	37.7±5.5	93.3±2.8
	DSI <0.2 or DSI >0.8									
	All features (1T, 1.5T, 3T)	14±4	12±5	5±2	0.81±0.10	83.3±7.4	98.9±4.2	67.7±13.6	59.0±17.4	99.4±2.3
	All features (>1T)	21±6	20±8	5±2	0.83±0.11	84.1±9.6	85.4±17.6	82.8±7.1	56.2±17.1	96.3±4.6
	DSI <0.3 or DSI >0.7									
	All features (1T, 1.5T, 3T)	37±6	34±7	10±3	0.84±0.06	80.7±6.0	89.6±11.4	71.8±9.4	47.8±10.9	96.8±3.0
NB	All features		291	63	0.78±0.05	68.2±5.2	54.6±11.4	81.8±5.8	40.4±8.2	89.3±2.2
	All featuresa		291	291	0.80±0.05	72.6±5.1	71.0±11.3	74.2±6.4	73.3±5.7	72.9±7.1
RF	All features		291	63	0.76±0.06	53.0±2.4	7.3±5.6	98.7±1.4	60.6±29.1	83.1±0.8
	All featuresa		291	291	0.76±0.06	67.6±6.4	49.6±11.2	85.6±6.0	77.1±9.5	63.4±5.1

For the extreme DSI values, n: number of patients in the test set having the DSI value in the given range; %: percentage of the patients in the test set (n=118) having the DSI value in the given range. [a] Number of progressive cases was increased by duplicating randomly selected progressive cases. Values are mean ± standard deviation over 10 iterations 3-fold CV.

Table 9. Classification performance of the classifiers evaluated by training with the ADC cohort and testing with the pooled DCN and BAR cohorts

Classi-fier	Variable	%	SSCD, n	PSCD, n	AUC	BACC	Sensitivity	Specificity	PPV	NPV
DSI	Demographics		232	88	0.63	57.8	61.4	54.3	33.8	78.8
	APOE		203	72	0.57	57.4	47.2	67.5	34.0	78.3
	Neuropsychology		232	88	0.69	63.9	63.6	64.2	40.3	82.3
	CSF		90	39	0.69	61.7	59.0	64.4	41.8	78.4
	MRI		100	42	0.77	67.4	73.8	61.0	44.3	84.7
	All features		232	88	0.72	65.1	68.2	62.1	40.5	83.7
	DSI <0.2 or DSI >0.8: All features	21	38	30	0.81	78.5	83.3	73.7	71.4	84.8
	DSI <0.3 or DSI >0.7: All features	43	94	50	0.79	74.2	76.0	72.3	59.4	85.0
NB	All features		232	88	0.71	67.2	51.1	83.2	53.6	81.8
	All features[a]		232	232	0.71	67.2	65.9	68.5	67.7	66.8
RF	All features		232	88	0.69	55.2	12.5	97.8	68.8	74.7
	All features[a]		232	232	0.66	62.1	47.4	76.7	67.1	59.3

For the extreme DSI values, n: number of patients having the DSI value in the given range; %: percentage of the patients having the DSI value in the given range. [a] Number of progressive cases was increased by duplicating randomly selected progressive cases.

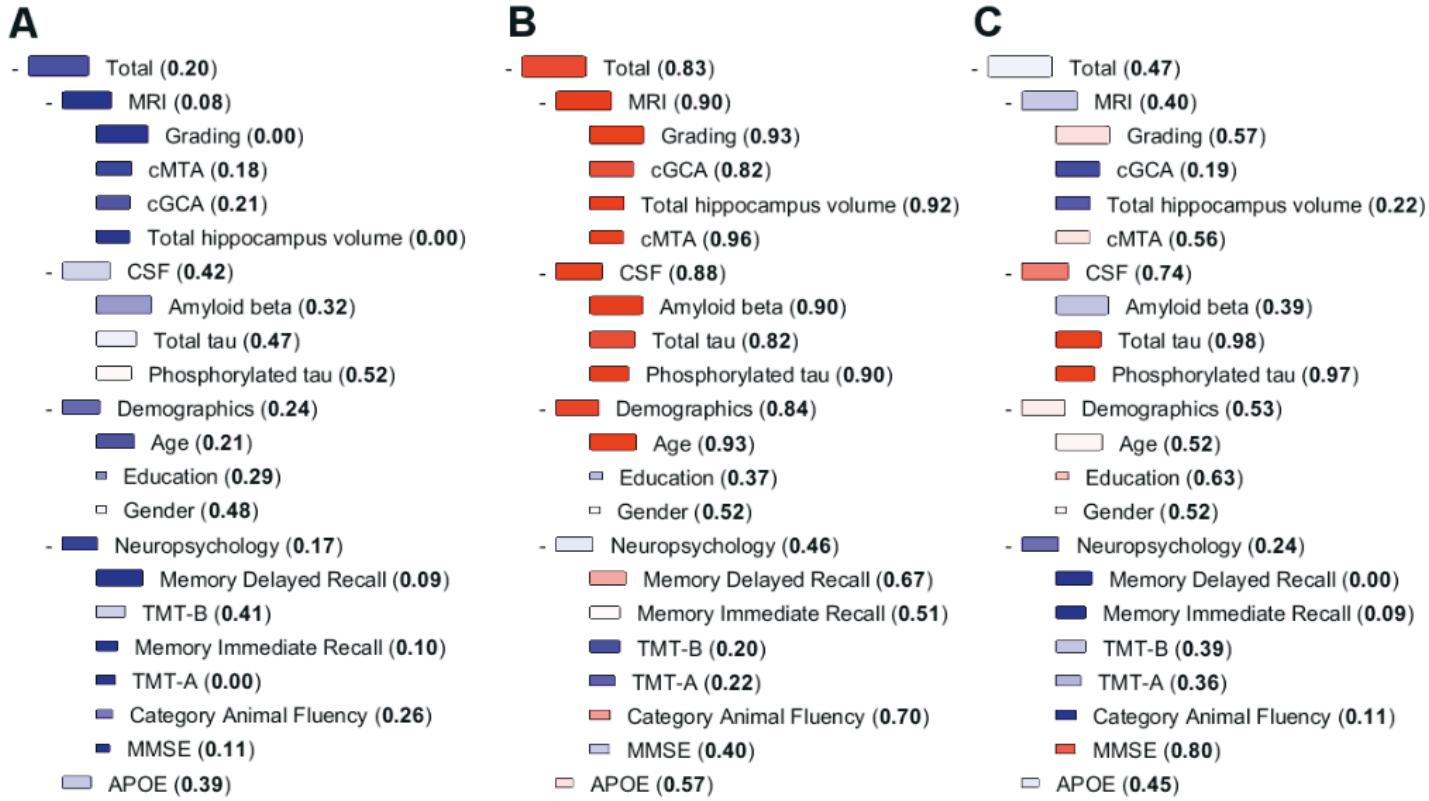


Figure 6. Examples of the Disease State Fingerprint visualizations: patients A and C remained stable and patient B progressed to MCI.

5.2 Monitoring progression of MCI (Study II)

Table 10 shows that longitudinal behaviour of the DSI values differed between the SMCI and PMCI groups. The PMCI cases had five times higher slopes and almost three times higher intercepts than the SMCI cases. The slopes of the SMCI cases were studied in more detail, and it was found that the SMCI group consisted of two subgroups: one group having lower slopes and another group having higher slopes which overlap with the slopes of the PMCI group (Figure 7). Table 11 shows classification performance of the regression parameters to differentiate the SMCI and PMCI cases. The best AUCs and classification accuracies were achieved when all data modalities were included, however, the increases were not always statistically significant.

The DSF visualization is an integral part of the DSI method. Figure 8 presents the DSF developed for longitudinal data. The data in the topmost panel is from a stable MCI patient. The DSI values and slope are low, indicating a stable disease status in this patient. The DSI values in the middle panel increase, indicating that the disease is progressing. MRI values of this patient show similarity to AD already from the beginning (boxes being red at all time points) and results from neuropsychology start pointing towards AD at the later phase (box colour changing from blue to red). This patient also remained stable during the 3-year follow-up. The data in the last panel is from a progressive MCI patient. The DSI values are high, and the slope is somewhat increasing, indicating that the markers of AD are present already at the beginning and the disease progresses over time.

Table 10. Regression parameters of longitudinal DSI values calculated using all variables

	SMCI	**PMCI**
Slope*	0.002 (0.000, 0.006) +	0.010 (0.005, 0.015) +
Intercept*	0.295 (0.139, 0.621) +	0.754 (0.626, 0.860) +
Number of points in regression	7 (7, 7)	5 (3, 5)

Values are median (25th percentile, 75th percentile). * statistically significant difference between SMCI and PMCI ($p<0.0005$, Mann-Whitney U test); + statistically significant difference from zero ($p<0.0005$, one-sample Wilcoxon signed rank test).

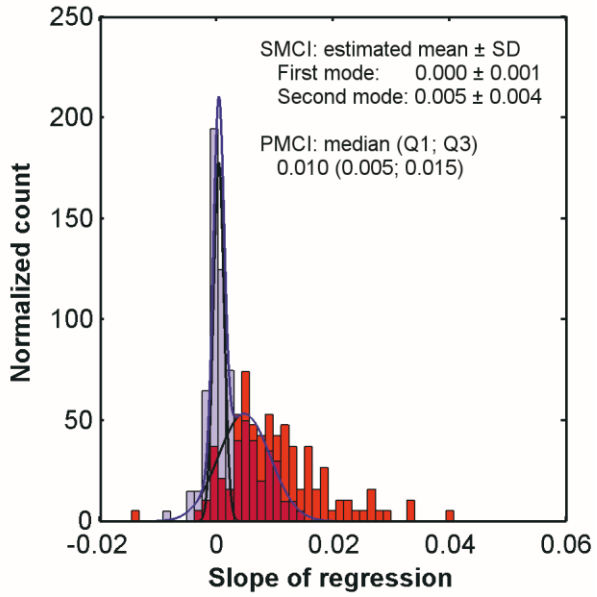


Figure 7. Histograms of the slopes for the SMCI (blue) and PMCI (red) cases. There appears to be two distinct subgroups in the SMCI group. Solid lines show a mixture distribution of two normal curves fitted to the slopes of the SMCI cases. The areas of the histograms are scaled to one. (SD: standard deviation, Q1: 25th quartile, Q3: 75th quartile).

Table 11. Classification performance of the regression parameters of the longitudinal DSI values obtained using different data modalities

	AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)
Slope				
All features	0.823	76.9±8.8	82.2±13.7	73.0±15.0
MMSE	0.771	71.8±7.6	55.5±15.5	86.5±5.5
ADAS-cog	0.768	68.7±10.2*	51.1±19.2	83.6±10.2
NeuroBat	0.766	69.2±5.8	60.2±13.2	76.9±15.3
MRI	0.710	66.8±8.1*	49.5±14.4	80.6±14.7
Intercept				
All features	0.808	74.6±8.7	75.1±17.4	74.4±12.2
MMSE	0.790	72.0±5.0	84.2±11.6	61.5±11.6
ADAS-cog	0.803	74.9±8.8	74.4±15.6	75.7±10.5
NeuroBat	0.793	66.9±6.1	74.4±21.7	61.0±14.0
MRI	0.696	60.4±8.9*	55.6±16.2	63.9±16.2

Values are mean ± standard deviation from the 10-fold CV, except for AUC. * statistically significant difference to the accuracy obtained using all features (Mann-Whitney U test or Student's t-test with Bonferroni correction).

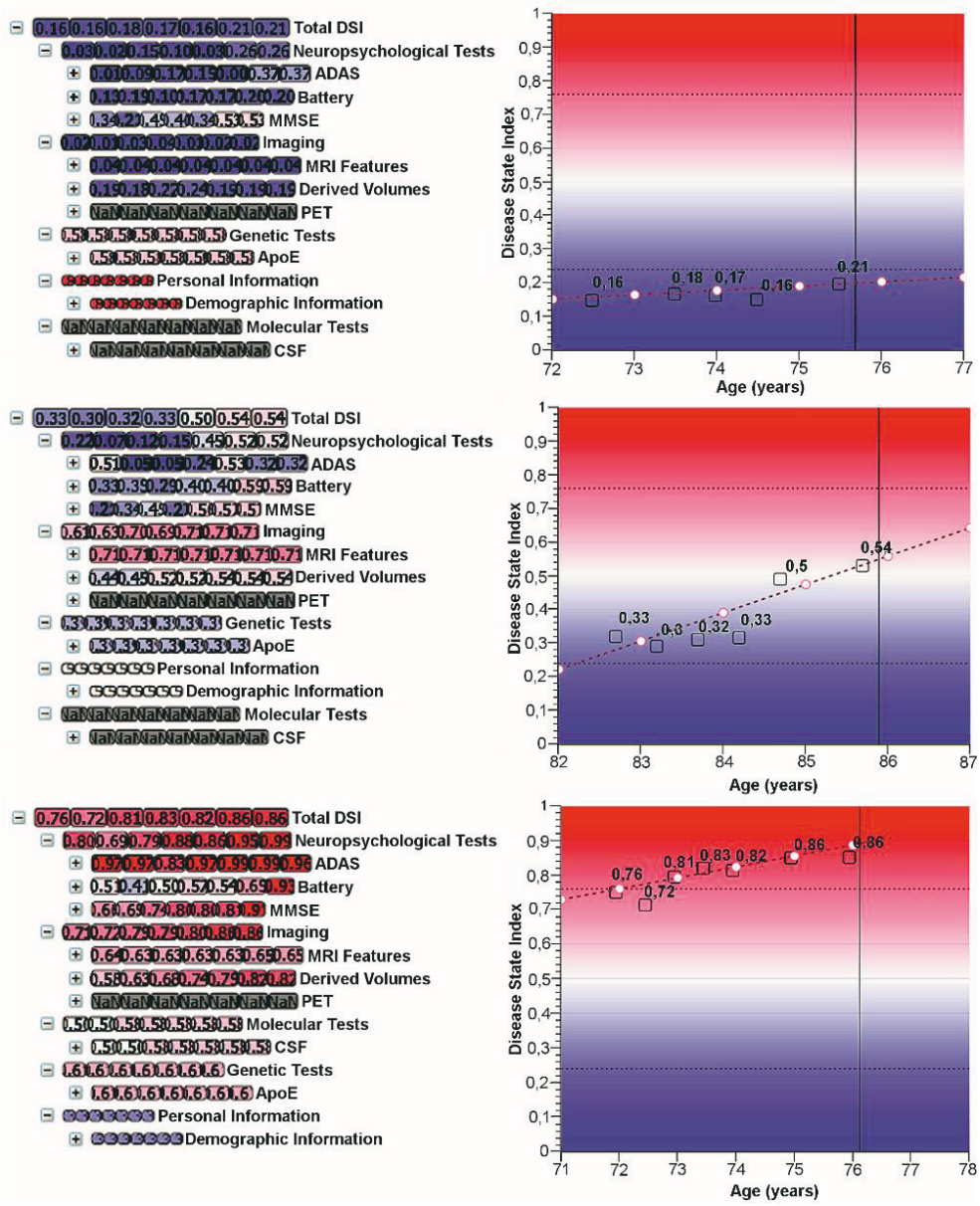


Figure 8. Visualizations for multimodal longitudinal data from three patients. Left panel: Longitudinal DSF showing on the rows the DSI values of individual tests at different time points. Right panel: The dashed line with white circles present regression line of the total DSI values over time for a patient, the black squares show the total DSI values of the patient, the horizontal line indicates a threshold were accuracy of 85% is achieved, and the vertical line shows the current age of the patient.

5.3 Prediction of hippocampal atrophy (Study III)

The results in Figure 9 and Table 12 show that the models including data from different modalities, like neuropsychological tests, MRI, CSF, and APOE, performed better than the models including only the MRI features. The full models had smaller RMSEs and higher correlation coefficients than the MRI-only models. Both models underestimated the real change at higher atrophy rate levels, the MRI-only models showing a greater underestimation (Figure 9).

Table 13 presents the performance of the models to predict dichotomized outcome, i.e., which of the subjects will have a fast rate of hippocampal atrophy. The accuracy was reasonable in all analyses (79-87%). The models including only MRI features had lower sensitivities than specificities. This is an expected result because the MRI-only models underestimated the higher atrophy rate levels.

In an attempt to improve the MRI-only models, the quadratic terms were added to the models. However, the changes in the performance were only minor and addition of the quadratic terms did not alleviate the problem of underestimation (results not shown here).

The models including only the MRI features generalized well to the independent AIBL cohort. The RMSEs were smaller in the validation in AIBL than in the cross-validation in ADNI 1. Also, the accuracy in prediction of fast atrophy rate was higher in AIBL than in ADNI 1. However, this was mostly due to the high specificities as the sensitivities were on the same level.

In all analyses, the LAD models performed better than the LMS models, which are more sensitive to outliers. The LMS models seemed to overpredict the low hippocampal atrophy rates and underpredict the high hippocampal atrophy rates. There were only minor differences between the models including both L1 and L2 norms (α=0.5) and the models including only the L1 norm (α=1.0). Thus, the detailed results for the L1 norm are not shown here.

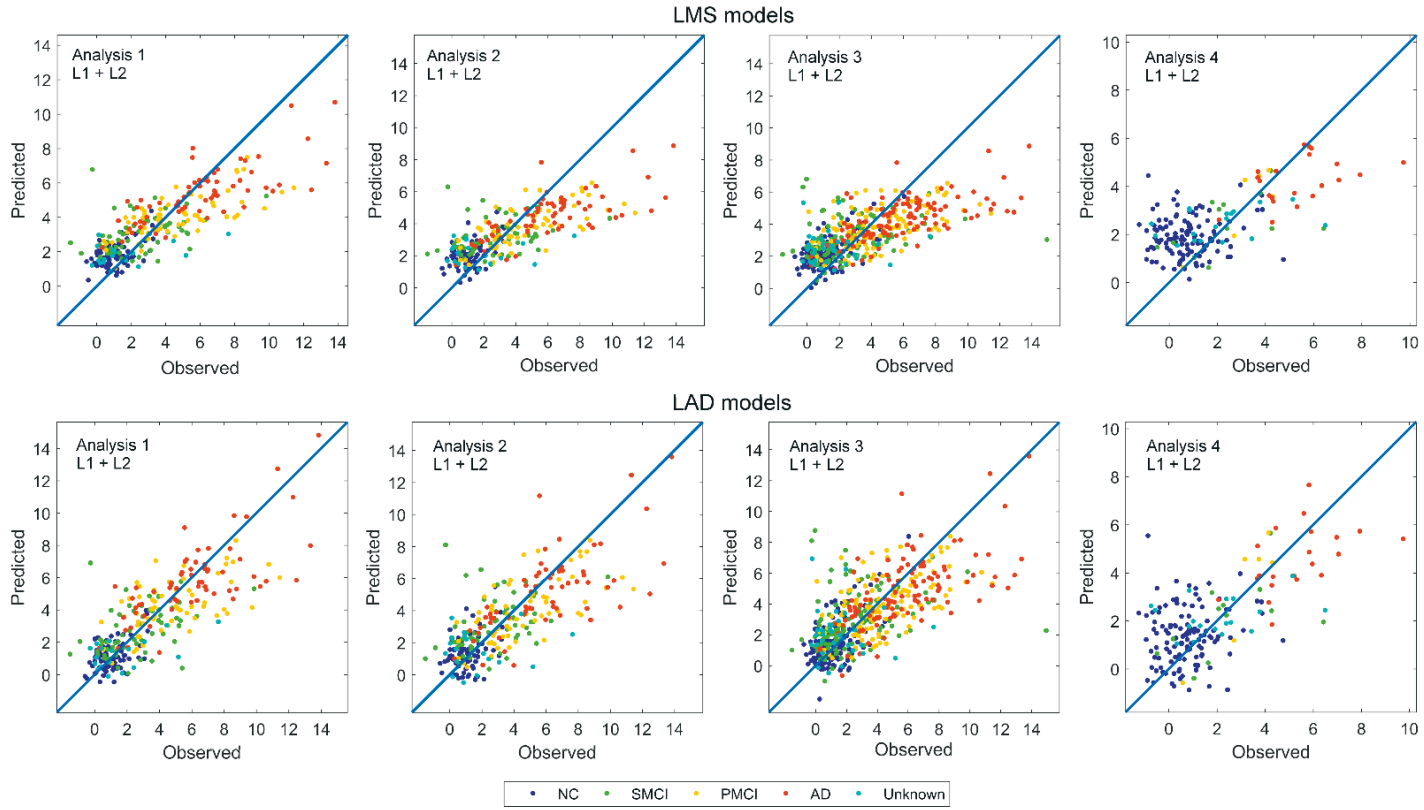


Figure 9. Observed and predicted annual hippocampal atrophy rates for the models with L1 and L2 norms in the regularization. Analysis 1 included the full set of variables, Analyses 2-4 included only the MRI features.

Table 12. Performance of the models to predict annual hippocampal atrophy rate (models with L1 and L2 norms in the regularization)

Model	#Features	RMSE (%)	Spearman rho
Analysis 1: full model + cross-validation with ADNI 1 (N=281)			
LMS	27 ± 2	1.79 ± 0.30**	0.78 ± 0.06
LAD	27 ± 2	1.76 ± 0.34*	0.76 ± 0.09
Analysis 2: MRI-only model + cross-validation with ADNI 1 (N=281)			
LMS	18 ± 4	2.06 ± 0.35	0.72 ± 0.09
LAD	18 ± 4	1.93 ± 0.44	0.72 ± 0.08
Analysis 3: MRI-only model + cross-validation with ADNI 1 (N=530)			
LMS	18 ± 4	2.11 ± 0.33	0.68 ± 0.07
LAD	18 ± 4	2.07 ± 0.38	0.68 ± 0.08
Analysis 4: MRI-only model + validation with AIBL (N=176)			
LMS	22	1.71 ± 0.06	0.71 ± 0.04
LAD	22	1.66 ± 0.07	0.71 ± 0.03

Values are mean ± standard deviation from the 10-fold CV (Analysis 1-3) or mean ± standard deviation over the repeated random selection of 20 HCs with all subjects in other diagnostic classes (to account for the imbalanced class distribution in AIBL, Analysis 4). * and **: p-value <0.01 and <0.05, respectively, when compared to the corresponding model in Analysis 2 (Wilcoxon signed rank test).

Table 13. Performance of the models to predict annual hippocampal atrophy rate, dichotomized to fast vs. slow (models with L1 and L2 norms in the regularization)

Model	Accuracy	Sensitivity	Specificity	PPV	NPV
Analysis 1: full model + cross-validation with ADNI 1 (N=281)					
LMS	0.84 ± 0.08	0.85 ± 0.14	0.84 ± 0.07	0.77 ± 0.10	0.89 ± 0.11
LAD	0.83 ± 0.08	0.84 ± 0.14	0.83 ± 0.07	0.75 ± 0.11	0.88 ± 0.11
Analysis 2: MRI-only model + cross-validation with ADNI 1 (N=281)					
LMS	0.82 ± 0.10	0.77 ± 0.13	0.85 ± 0.09	0.78 ± 0.13	0.84 ± 0.12
LAD	0.82 ± 0.08	0.78 ± 0.10	0.85 ± 0.10	0.79 ± 0.11	0.85 ± 0.08
Analysis 3: MRI-only model + cross-validation with ADNI 1 (N=530)					
LMS	0.79 ± 0.06	0.73 ± 0.12	0.82 ± 0.06	0.71 ± 0.09	0.83 ± 0.09
LAD	0.79 ± 0.04	0.73 ± 0.09	0.83 ± 0.08	0.72 ± 0.10	0.84 ± 0.06
Analysis 4: MRI-only model + validation with AIBL (N=176)					
LMS	0.87 ± 0.01	0.72 ± 0.01	0.97 ± 0.01	0.94 ± 0.02	0.84 ± 0.01
LAD	0.87 ± 0.01	0.72 ± 0.01	0.97 ± 0.01	0.93 ± 0.02	0.84 ± 0.02

Values are mean ± standard deviation from the stratified 10-fold CV (Analysis 1-3) or mean ± standard deviation over the repeated random selection of 20 HCs with all subjects in other diagnostic classes (to account for imbalanced class distribution in AIBL, Analysis 4).

The main difference between the L1 and the combination of L1 and L2 norms was that the use of L1 and L2 norms was less stringent regularization, thus, allowing more features to be included into the models. Features selected in all CV folds can be considered the most robust. In total, 19 features from all data modalities were selected in all folds when both L1 and L2 norms were used, whereas only 5 features were selected when the L1 norm was used. The L1 and L2 norms selected recalls of words (delayed recall (ball) from MMSE, word recall and delayed word recall from ADAS-cog, and trials 3 and 4 from RAVLT), orientation and modified total score from ADAS-cog; CSF Aβ and p-tau; presence of APOE ε4; and volumes of hippocampus, inferior lateral ventricles, and amygdala calculated with different methods. The L1 norm selected modified total score of ADAS-cog, presence of APOE ε4, volume of left hippocampus, volume of right inferior lateral ventricle, and volume of left hippocampus from TBM. Table 14 presents selected features when the MRI-only models were trained with ADNI 1 and tested with AIBL. The combination of L1 and L2 norms and the L1 norm alone selected almost the same features. In total, 12 features described volumes of hippocampus, inferior lateral ventricles, and amygdala.

Table 14. Selected MRI features when the models were trained with the ADNI 1 cohort and tested with the AIBL cohort (Analysis 4)

Variable	L1 + L2 (α = 0.5)	L1 (α = 1.0)
Volume of left hippocampus	x	x
Volume of right hippocampus	x	x
Volume of left amygdala	x	x
Volume of right amygdala	x	
Volume of left inferior lateral ventricle	x	x
Volume of right inferior lateral ventricle	x	x
Volume of left medial orbital gyrus	x	x
TBM: right amygdala	x	x
TBM: left amygdala	x	
TBM: left cerebellum exterior	x	x
TBM: left hippocampus	x	x
TBM: right inferior lateral ventricle	x	x
TBM: right entorhinal area	x	x
TBM: left middle temporal gyrus	x	x
TBM: right precuneus	x	x
TBM: right parahippocampal gyrus	x	x
VBM: global	x	x
VBM: left hippocampus	x	x
VBM: left inferior lateral ventricle	x	x
VBM: cerebellar vermal lobules VI-VII	x	x
VBM: left inferior temporal gyrus	x	x
VBM: left lingual gyrus	x	
VBM: right parahippocampal gyrus		x

5.4 Factors associated with mortality in AD (Study IV)

Table 15 presents associations of baseline variables with mortality, reported as HRs. All continuous variables were z-scored to allow comparison of results from different tests. Thus, the HRs should be interpreted in relation to standard deviations (i.e., the change in the risk of death if a value of a variable increases by one standard deviation). As male sex and older age were associated with an increased risk of mortality, the CPH models were adjusted for them. After the adjustment, older age, male sex, worse performance on MMSE, digit span backward, VAT naming, TMT-A, TMT-B, RAVLT immediate recall, and category fluency, as well as more severe MTA and GCA evaluated from MRI were associated with an increased risk of mortality. After additional adjustment for MMSE score and duration of complaints, associations remained between mortality and MMSE, digit span forward, TMT-A, TMT-B, and GCA. No associations were found between mortality and years of education, activities of daily living, duration of complaints, comorbidities, number of medications, and smoking (results not shown here). The optimal combination of variables, defined by forward selection, comprised age (HR 1.31, 95% CI 1.12-1.54, p=0.001), male sex (HR 1.67, 95% CI 1.26-2.21, p<0.001), digit span backward (HR 1.22, 95% CI 1.03-1.43, p=0.018), TMT-A (HR 1.22, 95% CI 1.06-1.41, p= 0.005), MTA (HR 1.18, 95% CI 1.01-1.38, p=0.038), and CSF p-tau (HR 1.15, 95% CI 1.00-1.32, p=0.058). Results on data without imputation were comparable with these results. Kaplan-Meier survival curves for these variables are shown in Figure 10.

Table 15. Associations between baseline variables and mortality

	Model 1: adjusted for age and sex		**Model 2: adjusted for age, sex, MMSE, and duration of complaints**	
	HR (95% CI)	p-value	HR (95% CI)	p-value
Demographics				
Age	1.29 (1.12-1.48)	<0.001	1.33 (1.15-1.53)	<0.001
Male sex	1.60 (1.22-2.11)	0.001	1.79 (1.35-2.37)	<0.001
Cognitive tests				
MMSE[a]	1.23 (1.07-1.42)	0.005	1.25 (1.08-1.44)	0.003
Digit span forward[a]	1.10 (0.95-1.26)	0.207	1.03 (0.89-1.20)	0.651
Digit span backward[a]	1.31 (1.13-1.52)	<0.001	1.24 (1.06-1.46)	0.009
VAT naming[a]	1.14 (1.01-1.30)	0.042	1.11 (0.97-1.27)	0.136
VAT memory[a]	1.07 (0.93-1.23)	0.360	1.02 (0.88-1.19)	0.790
TMT-A	1.29 (1.14-1.47)	<0.001	1.23 (1.08-1.41)	0.003
TMT-B	1.28 (1.13-1.45)	<0.001	1.21 (1.06-1.40)	0.005
RAVLT, immediate recall[a]	1.19 (1.02-1.38)	0.025	1.11 (0.95-1.30)	0.193
RAVLT, delayed recall[a]	0.96 (0.83-1.10)	0.507	0.90 (0.78-1.04)	0.154
Category fluency[a]	1.17 (1.01-1.36)	0.041	1.10 (0.94-1.29)	0.243
MRI				
MTA	1.18 (1.02-1.37)	0.030	1.15 (0.98-1.34)	0.081
PA	1.10 (0.95-1.28)	0.192	1.12 (0.96-1.29)	0.143
GCA	1.18 (1.01-1.37)	0.037	1.17 (1.00-1.36)	0.044
WMH	1.07 (0.92-1.25)	0.364	1.05 (0.90-1.22)	0.518
Lacunes present	1.10 (0.73-1.66)	0.634	1.17 (0.76-1.79)	0.485
Microbleed categories				
Microbleeds, 1-2	0.72 (0.43-1.19)	0.195	0.69 (0.42-1.16)	0.163
Microbleeds, ≥3	1.03 (0.76-1.40)	0.840	1.01 (0.74-1.37)	0.956
Infarcts present	1.15 (0.64-2.05)	0.641	1.11 (0.60-2.05)	0.727
CSF				
Aβ	1.02 (0.87-1.18)	0.850	0.99 (0.86-1.16)	0.943
total tau	1.09 (0.94-1.27)	0.275	1.07 (0.92-1.26)	0.369
p-tau	1.09 (0.94-1.26)	0.242	1.08 (0.93-1.26)	0.316
APOE ε4 carrier	1.11 (0.95-1.29)	0.204	1.09 (0.94-1.26)	0.278

All continuous variables were z-scored. [a] Scores were inverted (multiplied by -1) as originally lower scores indicated worse performance.

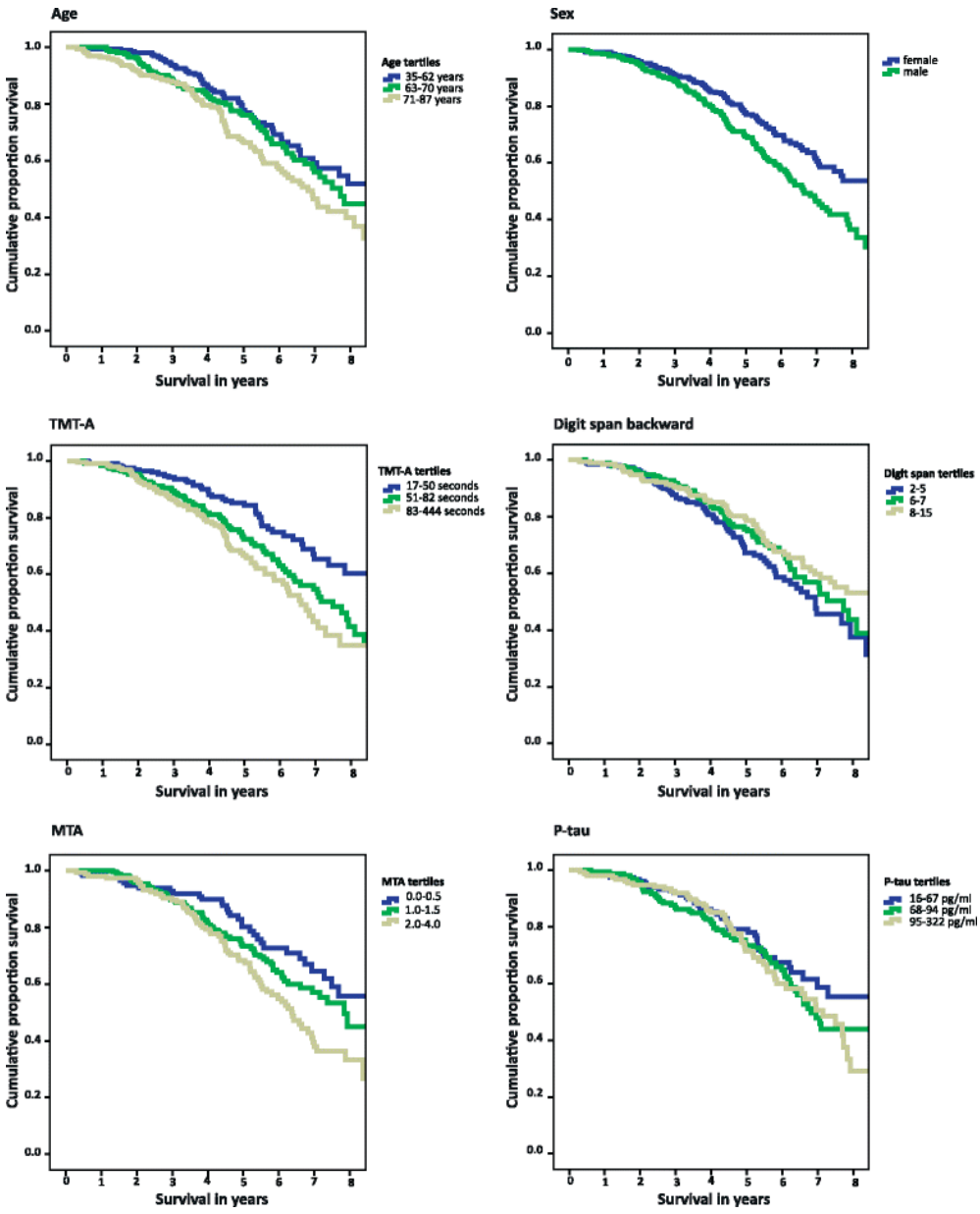


Figure 10. Kaplan-Meier curves for the optimal set of variables from forward selection. All except sex stratified in tertiles. The curves were plotted using raw data, without imputation.

6 DISCUSSION

6.1 Accomplishment of the objectives

The objective of this thesis was to develop and validate data-driven methods for predicting and monitoring progression of Alzheimer's disease at the different phases of the disease spectrum, starting from normal cognition and ending to death, using data from neuropsychological and cognitive tests, MRI, CSF, comorbidities, and APOE. Accomplishment of the specific objectives are discussed below.

The first objective was to develop a model for predicting who of the individuals with SCD are at risk for MCI or dementia.

This objective was accomplished by utilizing the DSI classifier and the large cohort of 674 subjects from three memory clinics in Europe for development and evaluation of the models. In the cohort, 22% of the SCD cases progressed to MCI or dementia during an average follow-up time of almost three years. The DSI classifier combining all available data modalities (demographics, APOE, neuropsychology, CSF, and MRI) obtained the best performance; the cross-validated BACC was 74%. All single- and multimodality classifiers had high NPV (>88%) and low PPV (<38%).

Almost half of the patients had an extreme DSI value (DSI < 0.3 or DSI > 0.7). In this subgroup, all performance measures were better when compared to the whole population. NPV was especially high (97%), meaning that progression in the patients with DSI < 0.3 is not probable and they could be reassured with good confidence. Whereas PPV was only modest (51%), meaning that the progressors cannot be identified as accurately, but a risk for progression is increased in patients with DSI > 0.7 and they might need more intensive follow-up.

When the models were validated with the independent test set, their performance decreased considerably (e.g., AUC decreased 11%). Additional analyses suggested that the DSI method itself did not worsen the generalizability of the models, instead, several differences between the cohorts caused the lower performance. In addition

to the DSI classifier, the NB and RF classifiers were used as reference methods and their performance was somewhat worse or corresponding.

The second objective was to develop methods for monitoring progression of disease over time in an MCI cohort in which some progressed to dementia due to AD and others did not.

This objective was accomplished by applying the DSI method to longitudinal data collected from an MCI cohort obtained from the ADNI 1 study. Longitudinal behaviour of the DSI values was assessed with linear regression of the DSI values over time. Longitudinal profiles of the DSI values differed between the SMCI and PMCI cases. The PMCI cases had five times higher slopes and almost three times higher intercepts than the SMCI cases, indicating that the PMCI cases had a more advanced disease state already at the beginning and a faster progression over time. In addition, two subgroups were found in the stable MCI group: one group with stable DSI values over time and another group with increasing DSI values. It was suggested that the group with lower slopes would have a truly stable disease status and the group with the higher slopes would progress to dementia if the follow-up time was longer. Other studies have also found that SMCI group is a heterogenous group with some subjects showing similarity to AD (Cui et al., 2011; Christos Davatzikos et al., 2011). This study also extended the DSF data visualization method for longitudinal data.

The third objective was to predict atrophy of hippocampus in a population consisting of subjects with normal cognition, MCI, and dementia due to AD.

This objective was accomplished by using the regularized linear regression and data from baseline to predict hippocampal atrophy over 24 months in subjects with normal cognition, MCI, or dementia due to AD. The multimodality models including neuropsychological and cognitive assessments, CSF, MRI features, and APOE performed better than the single-modality MRI models. The LAD models performed better than the LMS models, which is expected because the LAD models are less sensitive to outliers. The MRI models performed well when evaluated with an independent validation cohort.

All models underestimated the real change at the higher atrophy rate levels, the MRI models showing a greater underestimation. Addition of quadratic terms to the models did not alleviate the underestimation. This suggests that there might be other underlying factors which could explain the higher atrophy rates and were not in-

cluded in this study. E.g., cardiovascular disease, cardiac arrest, diabetes, hypertension, obstructive sleep apnoea, vitamin B_{12} deficiency, and head trauma have been associated with hippocampal atrophy (Fotuhi et al., 2012).

The most selected features for predicting hippocampal atrophy included word recall tasks, orientation and modified total score from the ADAS-cog, Aβ, p-tau, presence of APOE ε4, and MRI features describing hippocampus, inferior lateral ventricles, and amygdala. Many of the selected MRI features describe brain regions that are part of the medial temporal lobe which is affected in AD (Fox & Schott, 2004; Frisoni et al., 2010).

The fourth objective was to identify which of disease-related determinants are associated with mortality in patients with dementia due to AD.

This objective was accomplished by using the CPH model and a memory clinic cohort consisting of subjects with mild to moderate dementia due to AD. In this relatively young population (average age 67 years), 35% of the subjects died on average 5 years after the diagnosis. After adjustment for age and sex, older age, male sex, and worse scores on cognitive functioning, as well as more severe MTA and GCA were associated with an increased risk of mortality. An optimal combination of variables comprised age, sex, performance on digit span backward test and Trail Making Test A, MTA, and CSF p-tau.

6.2 Comparison to prior work

By the time of the publication (2018), Study I was among the first studies applying machine learning for predicting progression from SCD to MCI or dementia. The DSI models obtained comparable or even better performance than other developed models (Gómez-Ramírez et al., 2020; Guan et al., 2021; Liu et al., 2022, 2020; Yue et al., 2021). The other models were based on quite complex algorithms which may be difficult to understand for the clinicians. Especially, models based on neural networks have limited interpretability, whereas the DSI model and accompanied DSF visualisation are transparent and understandable for people who are not familiar with data science or machine learning, such as many clinicians. In fact, a group of clinicians was involved in the design of the DSI and DSF methods and interpretability was one of the key requirements. This may increase acceptability by the clinicians and thus increase uptake of these kinds of models into clinical practice. The strength

of Study I was the large sample size and validation of the models with the independent validation cohort. The proper validation highlighted the differences between the cohorts. Heterogeneity of the SCD cohorts is a known issue and the field is working towards harmonized research methods (Jessen et al., 2014; Molinuevo et al., 2017). Birkenbihl and colleagues (Birkenbihl et al., 2022) also observed clear differences in disease progression patterns between six well-known AD cohorts and found that data-driven models learn cohort-specific patterns, which limits generalizability of the models and underlines the need for validation with independent validation cohorts. Our cohorts were collected from the memory clinics where the data are incomplete and diverse, which is often the case in the real-life clinical practice. The developed models were able to provide reasonable predictions with these kinds of data. Validation with the independent cohorts and use of incomplete heterogenous data increased credibility of the models.

Study II investigated evolution of the DSI values over time. Other approaches for monitoring AD progression have also been proposed. Escudero and co-workers (Escudero et al., 2012) defined profiles of the disease and normality by applying k-means clustering to variables from the hypothetical model of AD progression (Jack et al., 2010, 2013). A subject's similarity to the profile of the disease was evaluated by calculating a so-called Bioindex, a continuous measure between [0, 1]. The Bioindex is based on normalized distances between the subject's data and the centroids of the profiles. To study progression of the Bioindices over time, Escudero and colleagues (Escudero et al., 2012) fitted a sigmoid function to each individuals' Bioindices. Similar to Study II, they also found that the PMCI cases had steeper progression towards AD than the SMCI cases. Other researchers have presented a different approach for modelling progression of AD. They fitted mathematical models with covariates to longitudinal ADAS-cog scores (Ito et al., 2011; Raket, 2020; Samtani et al., 2012). Samtani and colleagues (Samtani et al., 2012) restricted the analysis on patients with dementia due to AD and found that baseline disease status was affected by years since the onset of AD and baseline hippocampal and ventricular volumes. Progression rate was affected by baseline measures of age, total cholesterol, TMT-B score, current ADAS-cog score as well as APOE ε4 genotype. Ito and co-workers (Ito et al., 2011) utilized a wider population including HCs and patients with MCI or dementia due to AD. They found that rate of progression was influenced by the baseline disease severity, age, gender, and APOE ε4 genotype. Raket (Raket, 2020) utilized non-linear mixed effects models to model disease stage, baseline cognition, and the patients' individual changes in cognitive ability. Since the subjects entered the study at the different phases of the disease, the model was used to estimate a

common disease timeline, i.e., predicted disease time, which was used for aligning patient measurements. The predicted disease time was shown to better capture patterns of variation over time in other clinical and biomarker variables, e.g., in CDR-SB, FDG-PET, and hippocampal volume, than a conventional approach of creating separate trajectories for each diagnostic group. This model has later been extended to model disease progression using longitudinal measurements of three different clinical scores (ADAS-cog, MMSE, CDR-SB) simultaneously as outcomes (Kühnel et al., 2021).

Several studies in the field have predicted progression from one disease stage to another, e.g., from MCI to dementia due to AD. Study III used an alternative approach by predicting atrophy of the hippocampus over the 24 months. Hippocampal atrophy was chosen as an outcome because it is one of the well-known AD biomarkers. Other studies have used scores from the cognitive tests as outcomes (Bucholc et al., 2019; Huang et al., 2016; Weiner et al., 2017; D. Zhang et al., 2012a, 2012b; Zhou et al., 2013; Zhu et al., 2016). However, we decided to focus on hippocampal atrophy because it occurs relatively early on the disease spectrum, whereas changes in the cognitive test scores occur only in the later phases of the disease. In addition, the cognitive tests may include day-to-day variation due to various reasons, e.g., fatigue or alertness of the subject. Instead of using machine learning methods, other studies have utilized traditional statistical methods to search for factors that are associated with the hippocampal atrophy. They found similar variables as in Study III to be associated with the atrophy. Van de Pol and colleagues (van de Pol et al., 2007) found that older age, poorer general condition, APOE ε4, and baseline hippocampal volume were associated with higher rates of hippocampal atrophy in an MCI cohort with two years of follow-up. Henneman and co-workers (Henneman, Vrenken, et al., 2009) found CSF p-tau levels, baseline memory function, and visual rating of medial temporal lobe atrophy to be associated with hippocampal atrophy rate when corrected for age and gender. Similarly, Stricker and colleagues (Stricker et al., 2012) found baseline CSF Aβ and p-tau to be associated with hippocampal atrophy rate. A recent study on plasma biomarkers found that baseline plasma levels of tau phosphorylated at threonine 217 (p-tau217) and neurofilament light were independent predictors of longitudinal trajectories of hippocampal atrophy in a population consisting of HCs and subjects with SCD or MCI (Pereira et al., 2021). The study also found P-tau217 to be an independent predictor of temporal cortical thinning. Vuoksimaa and colleagues (Vuoksimaa et al., 2020) found that an MCI population with impaired performance on RAVLT had steeper descending trajectories of hippocampal and entorhinal cortical volumes over a period of 6-36 months than

HCs. Whereas MCI cases with normal performance on RAVLT had similar trajectories compared to HCs. The results were concordant when the analysis was restricted to the individuals with above average hippocampal or entorhinal cortex volumes at baseline. Also in Study III, RAVLT was among the most selected variables. Regarding prediction of atrophy in other brain regions than the hippocampus, participants of the TADPOLE machine learning challenge developed models for predicting future ADAS-cog scores and total ventricular volume (Marinescu et al., 2019). For the total ventricular volume, a model based on disease progression models and spline regression was the best with the mean absolute error of 0.41%. Interestingly, prediction of ADAS-cog was a more challenging problem as none of the methods was superior to the reference method based on linear mixed effects model.

Other studies on mortality mainly focused on older patients or more advanced dementia. The population in Study IV was relatively young and had mild to moderate dementia. Like earlier studies (Todd et al., 2013), Study IV found older age and male sex to be associated with mortality in AD. Cognitive impairment was evaluated using various instruments in other studies and the results were contradicting. Some studies found an association between mortality and rate of cognitive decline (Hui et al., 2003; Wilson et al., 2006). Study IV consistently found tests in the executive domain and, to a lesser degree, memory to be associated with mortality. The MRI results partly agreed and partly disagreed with findings from Hennemann and colleagues (Henneman, Sluimer, et al., 2009). In both studies, WMH were not associated with mortality in AD after adjustment for age and sex. Study IV found MTA and GCA to be associated with mortality, whereas Henneman and co-workers found no such associations in the AD population. However, they found GCA to be associated with mortality in younger patients (<68 years) when using the whole population (SCD, MCI, AD, other dementia, other diagnosis) and adjusting for sex and diagnosis. They also consistently found microbleeds to be associated with mortality, whereas Study IV found no such association. Nägga and colleagues (Nägga et al., 2014) also found MTA to associated with mortality in AD. Regarding CSF, other studies have found contradicting results (Boumenir et al., 2019; Degerman Gunnarsson et al., 2014; Nägga et al., 2014). Study IV follows this contradiction as CSF p-tau was included into the optimal combination of several variables but none of the CSF biomarkers were individually associated with mortality. Study IV found no associations between co-morbidities and mortality, whereas other studies found associations (van de Vorst et al., 2016). However, populations were older in those studies. Mank and co-workers (Mank et al., 2022) developed mortality prediction models for the AD continuum

from SCD to dementia due to AD. A multivariable model for SCD and MCI included age, sex, Neuropsychiatric Inventory, Charlson Comorbidity Index, GCA, MTA, CSF $A\beta_{42}$, and CSF p-tau. A multivariable model for patients with dementia due to AD included slightly different predictors: age, sex, MMSE, Neuropsychiatric Inventory, GCA, and WMH. Discriminative performance was higher in patients with SCD or MCI than in patients with dementia due to AD, which the authors discussed to possibly be due to a smaller and younger cohort with a lower burden of morbidity when compared to other studies. Deardorff and colleagues (Deardorff et al., 2022) also searched for an optimal combination of variables for prediction of mortality from the pool of demographics, behavioural and health factors, functional measures, and chronic conditions in community-dwelling adults with dementia. Unlike Study IV, their study did not include MRI or CSF measurements. Their final model included age, sex, body mass index, smoking status, activities of daily living dependency count, instrumental activities of daily living difficulty count, difficulty walking several blocks, participation in vigorous physical activity, and chronic conditions (cancer, heart disease, diabetes, lung disease). Finally, Study IV had the following strengths: 1) broad range of determinants were investigated; 2) all patients were assessed using the same harmonized diagnostic protocol and they received similar treatment and disease management; and 3) follow-up times were relatively long.

6.3 Impact of the research in its field

Impact for science: In terms of clinical impact, the developed prediction models increased understanding of the progression of AD; they revealed which factors were relevant for the progression in the different phases of the disease; and they found subgroups of patients with different disease progression profiles within the SMCI group. In terms of technical impact, this thesis showed in practice importance of the independent validation cohorts and heterogeneity of the SCD cohorts. Future studies need to consider these factors.

Impact for society: The developed prediction models may enable earlier diagnosis of AD so that the future treatment options can be started at the correct phase. Thus, patients would maintain their independence and ability to function longer. This would reduce costs to societies considerably due to reduced need for intensive nursing at the late stage of the disease. For example, it has been modelled that delaying the onset of AD by an average of 2 years would reduce worldwide prevalence of AD

in 2050 by 22.8 million cases (22%), of which 10.5 million cases (46%) would be in the late stage of the disease and need intensive nursing (Brookmeyer et al., 2007). Even a one-year delay in both disease onset and progression would decrease worldwide prevalence of AD by 9.2 million cases (9%), majority of these cases having the late-stage disease (8.7 million, 95%). Another study estimated that delaying the onset of the disease by one or five years would decrease population aged ≥70 years with AD in the United States in 2050 by 1.3 and 3.7 million cases (14% and 41%), respectively (Zissimopoulos et al., 2014). This study also estimated that without any disease-modifying treatments, total costs of AD to the economy of the United States would be $1.5 trillion in 2050 (in 2010 dollars). A delay of one or five years in the onset of the disease would reduce the costs by 15% and 40%, respectively. Another estimation showed that if a treatment delaying the onset of AD by five years became available in 2025, total costs of AD in the United State in 2050 would decrease by 33% from $1.101 trillion to $734 billion (in 2015 dollars) (Alzheimer's Association, 2015).

Impact for companies: The developed models may interest pharmaceutical companies as the models can potentially help in selecting suitable patient groups for clinical trials which investigate effects of disease-modifying therapies. E.g., the EMIF project, in which the Study III was conducted, included pharmaceutical companies like Janssen Pharmaceutica NV, Boehringer-Ingelheim International GmbH, Pfizer, and GlaxoSmithKline Research and Development Ltd. The models could also benefit medical technology companies, as they could be integrated as part of their existing products. One example of such collaboration is an integration of Quantib® Brain, which is a medical image analysis software, with a platform of GE Healthcare[2]. This kind of integration of new tools into existing and familiar products can improve acceptance of the new tools as the healthcare professionals do not need to learn yet another system. Finally, all four studies presented in this thesis were conducted in close collaboration with researchers, who founded a company called Combinostics Oy. Combinostics Oy offers a CDSS for diagnosis and management of different dementias. The DSI and DSF methods are key elements of their CDSS. Studies I and II in this thesis have thus contributed to the scientific foundation and credibility of this CDSS.

Scientifically proven effectiveness of CDSSs is very important for convincing investors, regulatory agencies, and customers, and finally for actual deployment in healthcare. Effectiveness of the CDSSs in AD can be evaluated with various metrics

[2] https://www.quantib.com/solutions/partnerships/quantib-brain, accessed on 4.12.2022

and the selection of the metrics depends on the precise use-case. The CDSSs need to demonstrate an adequate prediction performance when compared to diagnoses or prognoses given based on the current diagnostic guidelines. This reference has its known limitations, but it is the best available knowledge at the moment. Other metrics for the effectiveness can be evaluated in studies where clinicians assess their patients (or patients' data) with and without a CDSS. These metrics could include prediction accuracy of the clinicians, consistency of the predictions between the different clinicians, i.e., inter-rater agreement, clinicians' confidence in their own predictions, clinician's confidence in the prediction provided by the CDSS, time taken to review a patient's data, and a delay between the first visit to a memory clinic and a moment of receiving the most definitive diagnosis. In the case of selecting patients for clinical trials, the number of the participants required for a study could be used as the evaluation metric.

Impact for healthcare professionals: The DSI method and accompanying DSF visualizations have been integrated as part of the PredictAD and PredictND CDSSs. These CDSSs can help clinicians in the diagnostic work: it has been shown that the use of the PredictAD tool improved clinicians' confidence in making a diagnostic classification, accuracy of the diagnosis, and inter-rater agreement when compared to presenting all data on paper charts (Simonsen et al., 2013). The PredictND tool, which was developed for differential diagnostics of dementia, also improved clinicians' confidence in the diagnosis (Bruun et al., 2019). Furthermore, adding the tool to the diagnostic evaluation affected diagnostic decision making by changing the diagnosis of 13% of all the cases when compared to the diagnosis given without the tool (Bruun et al., 2019). However, the changes did not lead to statistically significant improvements in the prediction accuracy (Bruun et al., 2019). The DSI and the DSF methods may also enhance collaboration and communication between the healthcare professionals as well as communication between the healthcare professional and their patients resulting in improved care.

Impact for patients and their next of kin: It is important for the patients and their next of kin to receive correct diagnosis as early as possible to understand the reason for the symptoms, to obtain correct care and support, and to become accustomed living with the disease. Better communication with the healthcare professionals can improve the patients' understanding of the disease, which may motivate the patients for lifestyle changes if needed. In addition, the future treatments can be started at

the correct phase. All this can prolong the independent living with the disease which is important for the patients.

6.4 Limitations of the studies

The main limitation of all four studies was that the diagnoses of the patients were based on the clinical evaluation; and they were not confirmed with the 'ultimate diagnosis', which would be postmortem histological samples taken from the brain. However, collecting the histological samples would require a very long follow-up study because AD progresses slowly over the years and decades. The accuracy of the clinical AD diagnosis is known to be limited and the diagnoses are often delayed. Beach et al. (Beach et al., 2012) reported that sensitivity of the clinical AD diagnosis was 70.9–87.3% and specificity was 44.3–70.8%, depending on the used clinical and pathological criteria. In addition, Studies I-III were limited by the relatively short follow-up times, Study IV had a somewhat longer follow-up (4.3 years and 5.3 years among the patients who died and stayed alive, respectively). If the follow-up times were longer, the more patients might have progressed to more advanced stages and the diagnoses might be more accurate.

In Study I, an additional limitation was that we decided to include diagnostic tests and features which were already familiar to the clinicians or were found to be relevant in other studies. By extracting a larger number of MRI features and performing feature selection to find an optimal set of features might have improved the performance of the models.

In Study II, we utilized linear regression to model evolution of the DSI values over time because of its simplicity and paucity of the data. Some PMCI cases had only a few DSI values available because of their fast conversion to dementia due to AD and synchronization of the time stamps. More complex models might be more suitable for AD progression if the follow-up times are longer. E.g., Jack et al. (Jack et al., 2010, 2013) proposed AD biomarkers to have a sigmoidal shape; Caroli et al. (Caroli et al., 2010) found that for most biomarkers, except FDG-PET, the sigmoidal model fitted better than the linear model; and similarly Mouiha and Duchesne (Mouiha et al., 2012) found non-linear models to fit better than linear models to biomarkers other than FDG-PET. Another limitation is that we utilized data only from the ADNI 1 database. As was shown in Study I, it would be important to explore evolution of the DSI values in other cohorts as well.

In Study III, we were only able to assess the generalizability of the models including solely the MRI features because the independent validation cohort did not contain all the needed variables for the full models. In addition, the number of subjects with MCI or dementia due to AD was rather low in the validation cohort. It would be important to further validate the full and MRI models with additional cohorts including more subjects with MCI or dementia due to AD. This is especially important, because our models underestimated the real change at the higher atrophy rate levels. Furthermore, ADNI 1 had follow-up visits at months 12 and 24 whereas AIBL had them only at month 18. We had to scale the atrophy rate over 18 months in AIBL to correspond to atrophy rate over one year in ADNI 1. We used linear scaling which can be assumed to be valid due to such a short time interval when compared to the whole timeline of the disease. Finally, the hippocampus is also affected by conditions other than AD, e.g., normal ageing, diabetes, and sleep apnoea. Thus, if these kinds of models are taken into use in practice, e.g., when selecting suitable patients into the clinical trials of disease-modifying therapies, other AD biomarkers should also be taken into account simultaneously.

In Study IV, the generalizability of the results was limited by the population which was obtained from a tertiary memory clinic. Additionally, we also had limited information regarding medications. We only had medication use at the baseline and we did not have information about the use of cholinesterase inhibitors after the diagnosis. Some studies have shown cholinesterase inhibitors to increase survival, others have found no such effect, or the effect was found only in older patients (Lopez et al., 2009; Nordström et al., 2013; Rountree et al., 2012; Wattmo et al., 2015; Xu et al., 2021). Furthermore, we could not investigate the association between antipsychotics use and mortality because only a small number of subjects were on antipsychotics in our population.

6.5 Future work

Based on this thesis, there are various directions for the future work. In the following, the future work is described from a wide perspective on AD. Firstly, the DSI method and the DSF visualization have further development needs. The current version of the DSI method is purely for classification and it does not take time to event into account like the CPH model does. Other machine learning methods have already been adapted for survival data (Wang et al., 2019) and it would be interesting

to modify the DSI and DSF accordingly. The current version of the DSI also assumes that the feature values either increase or decrease as the pathology progresses. It does not take into account the possibility that both decreased and increased values can be pathological, e.g., too low or too high blood pressure values can indicate health problems. This development need was already mentioned in the thesis of Jussi Mattila (Mattila, 2014), but the work on this is still unfinished. Furthermore, the multiclass DSI has already been developed and there is an initial version of the multiclass DSF (Tolonen et al., 2018). Currently, there is work on-going on developing alternative visualisation options for the multiclass DSF that would be simple to interpret and understand.

Secondly, Study I showed that there were significant differences between the SCD cohorts which hampered generalization of the models. Research and clinical work would benefit from the harmonisation of the patient assessment protocols, diagnostic criteria, different tests and devices used. Understandably, this would be a lot of work and still some local variation might be needed from the practical viewpoint.

Thirdly, Study III has several development needs. The models should be further validated with additional cohorts with the full set of variables and a greater number of patients with MCI or dementia. As hippocampal atrophy is also affected by other diseases and conditions than AD, addition of variables describing co-morbidities or lifestyle might improve performance of the models, especially at the higher atrophy rate levels. Furthermore, as hippocampal atrophy is not specific only for AD, alternative outcomes for the prediction of state of AD pathology should be considered. Prediction of ATN biomarker status (Jack et al., 2016), which takes amyloid load, tau pathology, and neurodegeneration simultaneously into account, would be an interesting approach.

Fourthly, development of machine learning methods requires data and quality of data (e.g., volume, representativeness, goodness of features) is a key determinant for the goodness of the models. As AD develops gradually over the years, collecting new cohorts for answering new research questions requires time and money. Biobanks, hospital data lakes, and electronic health records provide an alternative data source to prospective research cohorts. An example study on the use of the real-world data is the Pharmacogenomics of Antithrombotic Drugs (PreMed) study which combined data from the Finnish biobanks and various health registries, including demographic, genomic, health encounter, drug dispensation, patient record, and laboratory data (Lähteenmäki et al., 2021, 2022; Vuorinen et al., 2021). Similar approach could be utilized to study AD and dementia related research questions (e.g., other

predictors of AD, effects of co-morbidities and medications, genetics, use of healthcare resources). These data sources could also serve as independent validation data sets. Research cohorts typically have well-defined study protocols with certain eligibility criteria, fixed follow-up intervals, and selected patient assessments. Real-world data from the clinics is much more heterogenous and often incomplete. If the machine learning models are taken into use in a real healthcare environment, it would be important to test their performance and generalizability with not so optimal and heterogenous data.

Fifthly, lots of different machine learning methods have been applied to AD to predict various outcomes, however, only few of them have been deployed in the healthcare environment in practice, e.g., tools from the companies like Combinostics Oy[3], icometrix[4], IXICO plc[5], and Quantib B. V.[6]. The questions are how to bridge the gap between the machine learning research in AD and actual implementation and deployment of the methods in clinics; how to integrate machine learning methods as part of the hospital information systems and as part of the professionals' workflows; and what kinds of barriers and enablers there are for the deployment.

Sixthly, there is a need for low-cost, less invasive, and accurate screening tools for early detection and monitoring of AD. Currently used and well-established cognitive tests, such as MMSE, are less effective in detecting AD in its earliest stages. Lately, advances in development of sensitive assays have paved the way for blood-based biomarkers to become a reality in detection and monitoring of AD (Teunissen et al., 2022). However, there are still various clinical, technical, regulatory, and ethical issues to be solved before the blood-based biomarkers are used in practice, e.g., performance in the real-world and in different settings; cost-effectiveness; integration in clinical guidelines, in-vitro-diagnostic assay development, validation, and certification (Teunissen et al., 2022). Another interesting field is digital biomarkers for detection and monitoring of AD. Digital biomarkers are extracted from data collected with mobile, wearable, and environmental devices, such as smart phones, rings, watches; sensor patches; sensors integrated to the clothes; cameras and infra-red motion sensors. These biomarkers have been less studied than traditional cognitive and imaging biomarkers. There is a need for larger sample sizes, inclusion of patients from the different phases of the disease spectrum, longitudinal data, and validation of the methods with independent test sets.

[3] https://www.combinostics.com/, accessed on 4.12.2022
[4] https://www.icometrix.com/, accessed on 4.12.2022
[5] https://www.ixico.com/, accessed on 4.12.2022
[6] https://www.quantib.com/, accessed on 4.12.2022

Finally, as lifestyle plays an important role in preventing and delaying AD, an evident question is how to motivate people to acquire and maintain healthy lifestyle throughout their lives? This is not an easy question because, after a vast amount of research on lifestyle interventions for other diseases and information sharing for the public, increasing obesity rates, lack of physical activity, and sedentarism are still global health issues. Multidomain lifestyle interventions for AD have shown their effectiveness (Kivipelto et al., 2018; Ngandu et al., 2015), however, providing extensive face-to-face lifestyle counselling for individuals or groups through healthcare professionals becomes easily too expensive when targeted for larger populations. Digital tools offer potentially cost-effective and scalable ways to deliver lifestyle interventions and a few tools have already been developed for AD and related dementias (Bott et al., 2019). Typically, digital lifestyle interventions suffer from high attrition and drop-out rates (Eysenbach, 2005). Thus, it is important to study factors affecting adherence to the digital lifestyle interventions in order to design effective interventions and to select the right intervention for the right people. Both statistical and machine learning methods can be utilized for this task. Most of the adherence studies on mobile health apps targeting prevention and management of noncommunicable diseases had a pilot character with short study durations (Jakob et al., 2022). In addition, there is a scarcity of studies regarding adherence to digital lifestyle interventions for dementia prevention (Jakob et al., 2022).

7 CONCLUSIONS

This thesis presented four studies on development and validation of the data-driven methods for predicting and monitoring progression of AD at the different phases of the disease spectrum. The data were obtained from publicly available databases (US, Australia) and from non-public cohorts (the Netherlands, Germany, Spain). The data included neuropsychological and cognitive tests, MRI, CSF, comorbidities, and APOE. The main findings of the thesis can be summarized as follows:

- The developed models provide valid means to predict and follow progression of AD from the mildest stages to the more advanced stages.

- The models based on several different data modalities obtained better prediction performance than the models based on the single data modality.

- It is extremely important to evaluate generalizability of the models with independent validation cohorts.

- Harmonization of patient assessment methods, diagnostic criteria, tests, and devices may be needed if these methods are introduced to different settings and countries.

Research contribution presented in the thesis is scientifically novel as 1) by the time of its publication (2018), Study I was among the first studies utilizing machine learning for predicting progression of the disease in the SCD population; 2) Study II extended the DSF visualization from a single time point to longitudinal data and investigated evolution of the DSI values over time; 3) Study III utilized hippocampal atrophy, an established biomarker of AD pathology, as a continuous outcome whereas most other studies had focussed on continuous scores from cognitive and neuropsychological tests; 4) Study IV extended the knowledge on determinants of mortality in AD in younger patients with mild to moderate AD, for whom there is

paucity of data. The developed models may aid in the early diagnosis of AD, selection of suitable patients for trials investigating disease-modifying therapies, and follow-up of the disease progression over time.

REFERENCES

Acock, A. C. (2005). Working With Missing Values. *Journal of Marriage and Family*, *67*(4), 1012–1028. https://doi.org/10.1111/J.1741-3737.2005.00191.X

Aisen, P. S., Cummings, J., Jack, C. R., Morris, J. C., Sperling, R., Frölich, L., Jones, R. W., Dowsett, S. A., Matthews, B. R., Raskin, J., Scheltens, P., & Dubois, B. (2017). On the path to 2025: understanding the Alzheimer's disease continuum. *Alzheimer's Research & Therapy*, *9*(1), 60. https://doi.org/10.1186/s13195-017-0283-5

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., Gamst, A., Holtzman, D. M., Jagust, W. J., Petersen, R. C., Snyder, P. J., Carrillo, M. C., Thies, B., & Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 270–279. https://doi.org/10.1016/j.jalz.2011.03.008

Alzheimer's Association. (2015). *Changing the Trajectory of Alzheimer's Disease: How a Treatment by 2025 Saves Lives and Dollars*. https://www.alz.org/media/Documents/changing-the-trajectory-r.pdf

Alzheimer's Association. (2018). 2018 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *14*(3), 367–429. https://doi.org/10.1016/J.JALZ.2018.02.001

Alzheimer's Association. (2021). 2021 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, *17*, 327–406. https://doi.org/10.1002/alz.12328

Antila, K., Lötjönen, J., Thurfjell, L., Laine, J., Massimini, M., Rueckert, D., Zubarev, R. A., Orešič, M., van Gils, M., Mattila, J., Hviid Simonsen, A., Waldemar, G., & Soininen, H. (2013). The PredictAD project: development of novel biomarkers and analysis software for early diagnosis of the Alzheimer's disease. *Interface Focus*, *3*(2), 20120072. https://doi.org/10.1098/rsfs.2012.0072

Asanomi, Y., Shigemizu, D., Akiyama, S., Sakurai, T., Ozaki, K., Ochiya, T., & Niida, S. (2021). Dementia subtype prediction models constructed by penalized regression methods for multiclass classification using serum microRNA expression data. *Scientific Reports 2021 11:1*, *11*(1), 1–8. https://doi.org/10.1038/s41598-021-00424-1

Ashburner, J., & Friston, K. J. (2000). Voxel-based morphometry—the methods. *NeuroImage*, *11*, 805–821. https://doi.org/10.1006/nimg.2000.0582

Bang, J., Spina, S., & Miller, B. L. (2015). Frontotemporal dementia. *The Lancet*, *386*(10004), 1672–1682. https://doi.org/10.1016/S0140-6736(15)00461-4

Battista, P., Salvatore, C., Berlingeri, M., Cerasa, A., & Castiglioni, I. (2020). Artificial intelligence and

neuropsychological measures: The case of Alzheimer's disease. *Neuroscience & Biobehavioral Reviews*, *114*, 211–228. https://doi.org/10.1016/J.NEUBIOREV.2020.04.026

Beach, T. G., Monsell, S. E., Phillips, L. E., & Kukull, W. (2012). Accuracy of the Clinical Diagnosis of Alzheimer Disease at National Institute on Aging Alzheimer's Disease Centers, 2005–2010. *Journal of Neuropathology and Experimental Neurology*, *71*(4), 266. https://doi.org/10.1097/NEN.0B013E31824B211B

Bibl, M., Esselmann, H., & Wiltfang, J. (2012). Neurochemical biomarkers in Alzheimer's disease and related disorders. *Therapeutic Advances in Neurological Disorders*, *5*(6), 335. https://doi.org/10.1177/1756285612455367

Birkenbihl, C., Salimi, Y., & Fröhlich, H. (2022). Unraveling the heterogeneity in Alzheimer's disease progression across multiple cohorts and the implications for data-driven disease modeling. *Alzheimer's & Dementia*, *18*(2), 251–261. https://doi.org/10.1002/ALZ.12387

Bond, J., Stave, C., Sganga, A., O'Connell, B., & Stanley, R. L. (2005). Inequalities in dementia care across Europe: key findings of the Facing Dementia Survey. *International Journal of Clinical Practice*, *59*(Suppl. 146), 8–14. https://doi.org/10.1111/J.1368-504X.2005.00480.X

Bott, N. T., Hall, A., Madero, E. N., Glenn, J. M., Fuseya, N., Gills, J. L., & Gray, M. (2019). Face-to-Face and Digital Multidomain Lifestyle Interventions to Enhance Cognitive Reserve and Reduce Risk of Alzheimer's Disease and Related Dementias: A Review of Completed and Prospective Studies. *Nutrients*, *11*(9), 2258. https://doi.org/10.3390/NU11092258

Boumenir, A., Cognat, E., Sabia, S., Hourregue, C., Lilamand, M., Dugravot, A., Bouaziz-Amar, E., Laplanche, J. L., Hugon, J., Singh-Manoux, A., Paquet, C., & Dumurgier, J. (2019). CSF level of β-amyloid peptide predicts mortality in Alzheimer's disease. *Alzheimer's Research & Therapy*, *11*, 29. https://doi.org/10.1186/S13195-019-0481-4

Bradburn, M. J., Clark, T. G., Love, S. B., & Altman, D. G. (2003). Survival analysis part II: multivariate data analysis--an introduction to concepts and methods. *British Journal of Cancer*, *89*(3), 431–436. https://doi.org/10.1038/sj.bjc.6601119

Bradford, A., Kunik, M. E., Schulz, P., Williams, S. P., & Singh, H. (2009). Missed and Delayed Diagnosis of Dementia in Primary Care: Prevalence and Contributing Factors Inquiry to Improve Outpatient Safety Through Effective Electronic Communication, both at the. *Alzheimer Disease & Associated Disorders*, *23*(4), 306–314. https://doi.org/10.1097/WAD.0b013e3181a6bebc

Bratić, B., Kurbalija, V., Ivanović, M., Oder, I., & Bosnić, Z. (2018). Machine Learning for Predicting Cognitive Diseases: Methods, Data Sources and Risk Factors. *Journal of Medical Systems*, *42*(12), 243. https://doi.org/10.1007/S10916-018-1071-X

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Bron, E. E., Smits, M., Papma, J. M., Steketee, R. M. E., Meijboom, R., de Groot, M., van Swieten, J. C., Niessen, W. J., & Klein, S. (2017). Multiparametric computer-aided differential diagnosis of Alzheimer's disease and frontotemporal dementia using structural and advanced MRI. *European*

Radiology, *27*(8), 3372–3382. https://doi.org/10.1007/S00330-016-4691-X

Brookmeyer, R., Johnson, E., Ziegler-Graham, K., & Arrighi, H. M. (2007). Forecasting the global burden of Alzheimer's disease. *Alzheimer's & Dementia*, *3*, 186–191. https://doi.org/10.1016/j.jalz.2007.04.381

Bruun, M., Frederiksen, K. S., Rhodius-Meester, H. F. M., Baroni, M., Gjerum, L., Koikkalainen, J., Urhemaa, T., Tolonen, A., van Gils, M., Tong, T., Guerrero, R., Rueckert, D., Dyremose, N., Andersen, B. B., Simonsen, A. H., Lemstra, A., Hallikainen, M., Kurl, S., Herukka, S.-K., … Hasselbalch, S. G. (2019). Impact of a Clinical Decision Support Tool on Dementia Diagnostics in Memory Clinics: The PredictND Validation Study. *Current Alzheimer Research*, *16*(2), 91–101. https://doi.org/10.2174/1567205016666190103152425

Bucholc, M., Ding, X., Wang, H., Glass, D. H., Wang, H., Prasad, G., Maguire, L. M., Bjourson, A. J., McClean, P. L., Todd, S., Finn, D. P., Wong-Lin, K., & the Alzheimer's Disease Neuroimaging Initiative. (2019). A practical computerized decision support system for predicting the severity of Alzheimer's disease of an individual. *Expert Systems with Applications*, *130*, 157–171. https://doi.org/10.1016/J.ESWA.2019.04.022

Canu, E., Agosta, F., Mandic-Stojmenovic, G., Stojković, T., Stefanova, E., Inuggi, A., Imperiale, F., Copetti, M., Kostic, V. S., & Filippi, M. (2017). Multiparametric MRI to distinguish early onset Alzheimer's disease and behavioural variant of frontotemporal dementia. *NeuroImage: Clinical*, *15*, 428–438. https://doi.org/10.1016/J.NICL.2017.05.018

Caroli, A., Frisoni, G. B., & Initiative, T. A. D. N. (2010). The dynamics of Alzheimer's disease biomarkers in the Alzheimer's Disease Neuroimaging Initiative cohort. *Neurobiology of Aging*, *31*(8), 1263. https://doi.org/10.1016/J.NEUROBIOLAGING.2010.04.024

Cattel, C., Gambassi, G., Sgadari, A., Zuccalà, G., Carbonin, P., & Bernabei, R. (2000). Correlates of delayed referral for the diagnosis of dementia in an outpatient population. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, *55*(2), M98-102. https://doi.org/10.1093/GERONA/55.2.M98

Cavedoni, S., Chirico, A., Pedroli, E., Cipresso, P., & Riva, G. (2020). Digital Biomarkers for the Early Detection of Mild Cognitive Impairment: Artificial Intelligence Meets Virtual Reality. *Frontiers in Human Neuroscience*, *14*, 245. https://doi.org/10.3389/FNHUM.2020.00245

Cui, Y., Liu, B., Luo, S., Zhen, X., Fan, M., Liu, T., Zhu, W., Park, M., Jiang, T., Jin, J. S., & Alzheimer's Disease Neuroimaging Initiative. (2011). Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors. *PloS One*, *6*(7), e21896. https://doi.org/10.1371/journal.pone.0021896

Cummings, J., Aisen, P. S., Dubois, B., Frölich, L., Jack, C. R., Jones, R. W., Morris, J. C., Raskin, J., Dowsett, S. A., & Scheltens, P. (2016). Drug development in Alzheimer's disease: The path to 2025. *Alzheimer's Research and Therapy*, *8*(1), 39. https://doi.org/10.1186/s13195-016-0207-9

Cummings, J., Lee, G., Ritter, A., Sabbagh, M., & Zhong, K. (2020). Alzheimer's disease drug development pipeline: 2020. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *6*(1). https://doi.org/10.1002/trc2.12050

Das, D., Ito, J., Kadowaki, T., & Tsuda, K. (2019). An interpretable machine learning model for diagnosis of Alzheimer's disease. *PeerJ*, *7*, e6543. https://doi.org/10.7717/PEERJ.6543

Dasgupta, M., & Mishra, S. K. (2004). Least Absolute Deviation Estimation of Linear Econometric Models: A Literature Review. *SSRN*. https://doi.org/10.2139/SSRN.552502

Davatzikos, C., Resnick, S. M., Wu, X., Parmpi, P., & Clark, C. M. (2008). Individual patient diagnosis of AD and FTD via high-dimensional pattern classification of MRI. *NeuroImage*, *41*(4), 1220–1227. https://doi.org/10.1016/J.NEUROIMAGE.2008.03.050

Davatzikos, Christos, Bhatt, P., Shaw, L. M., Batmanghelich, K. N., & Trojanowski, J. Q. (2011). Prediction of MCI to AD conversion, via MRI, CSF biomarkers, pattern classification. *Neurobiology of Aging*, *32*(12), 2322.e19. https://doi.org/10.1016/J.NEUROBIOLAGING.2010.05.023

de Miranda, L. F. J. R., de Oliveira Matoso, R., Rodrigues, M. V., de Lima, T. O. L., Nascimento, A. F., Carvalho, F. C., de Melo Moreira, D. R., Fernandes, J. C., de Paula, J. J., Magno, L. A. V., Caramelli, P., & de Moraes, E. N. (2011). Factors influencing possible delay in the diagnosis of Alzheimer's disease: Findings from a tertiary Public University Hospital. *Dementia & Neuropsychologia*, *5*(4), 328–331. https://doi.org/10.1590/S1980-57642011DN05040011

Deardorff, W. J., Barnes, D. E., Jeon, S. Y., Boscardin, W. J., Langa, K. M., Covinsky, K. E., Mitchell, S. L., Whitlock, E. L., Smith, A. K., & Lee, S. J. (2022). Development and External Validation of a Mortality Prediction Model for Community-Dwelling Older Adults With Dementia. *JAMA Internal Medicine*, *182*(11), 1161–1170. https://doi.org/10.1001/jamainternmed.2022.4326

Degerman Gunnarsson, M., Lannfelt, L., Ingelsson, M., Basun, H., & Kilander, L. (2014). High Tau Levels in Cerebrospinal Fluid Predict Rapid Decline and Increased Dementia Mortality in Alzheimer's Disease. *Dementia and Geriatric Cognitive Disorders*, *37*(3–4), 196–206. https://doi.org/10.1159/000355556

Dennis, E. L., & Thompson, P. M. (2014). Functional Brain Connectivity Using fMRI in Aging and Alzheimer's Disease. *Neuropsychology Review*, *24*(1), 49–62. https://doi.org/10.1007/s11065-014-9249-6

Diehl, J., Monsch, A. U., Aebi, C., Wagenpfeil, S., Krapp, S., Grimmer, T., Seeley, W., Förstl, H., & Kurz, A. (2005). Frontotemporal dementia, semantic dementia, and Alzheimer's disease: the contribution of standard neuropsychological tests to differential diagnosis. *Journal of Geriatric Psychiatry and Neurology*, *18*(1), 39–44. https://doi.org/10.1177/0891988704272309

Dielman, T. E. (2005). Least absolute value regression: recent contributions. *Journal of Statistical Computation and Simulation*, *75*(4), 263–286. https://doi.org/10.1080/0094965042000223680

Dong, Q., Zhang, J., Li, Q., Wang, J., Leporé, N., Thompson, P. M., Caselli, R. J., Ye, J., & Wang, Y. (2020). Integrating Convolutional Neural Networks and Multi-task Dictionary Learning for Cognitive Decline Prediction with Longitudinal Images. *Journal of Alzheimer's Disease*, *75*(3), 971–992. https://doi.org/10.3233/JAD-190973

Duara, R., Barker, W., Loewenstein, D., & Bain, L. (2009). The basis for disease-modifying treatments for Alzheimer's disease: The Sixth Annual Mild Cognitive Impairment Symposium. *Alzheimer's &*

Dementia, *5*(1), 66–74. https://doi.org/10.1016/j.jalz.2008.10.006

Dubois, B., Feldman, H. H., Jacova, C., Cummings, J. L., DeKosky, S. T., Barberger-Gateau, P., Delacourte, A., Frisoni, G., Fox, N. C., Galasko, D., Gauthier, S., Hampel, H., Jicha, G. A., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., … Scheltens, P. (2010). Revising the definition of Alzheimer's disease: a new lexicon. *The Lancet Neurology*, *9*(11), 1118–1127. https://doi.org/10.1016/S1474-4422(10)70223-4

Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P. J., & Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria. *The Lancet Neurology*, *6*(8), 734–746. https://doi.org/10.1016/S1474-4422(07)70178-3

Dubois, B., Feldman, H. H., Jacova, C., Hampel, H., Molinuevo, J. L., Blennow, K., DeKosky, S. T., Gauthier, S., Selkoe, D., Bateman, R., Cappa, S., Crutch, S., Engelborghs, S., Frisoni, G. B., Fox, N. C., Galasko, D., Habert, M.-O., Jicha, G. A., Nordberg, A., … Cummings, J. L. (2014). Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *The Lancet Neurology*, *13*(6), 614–629. https://doi.org/10.1016/S1474-4422(14)70090-0

Dubois, B., Hampel, H., Feldman, H. H., Scheltens, P., Aisen, P., Andrieu, S., Bakardjian, H., Benali, H., Bertram, L., Blennow, K., Broich, K., Cavedo, E., Crutch, S., Dartigues, J.-F., Duyckaerts, C., Epelbaum, S., Frisoni, G. B., Gauthier, S., Genthon, R., … Jack, C. R. (2016). Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia*, *12*, 292–323. https://doi.org/10.1016/j.jalz.2016.02.002

Dubois, B., Padovani, A., Scheltens, P., Rossi, A., & Dell'agnello, G. (2016). Timely Diagnosis for Alzheimer's Disease: A Literature Review on Benefits and Challenges. *Journal of Alzheimer's Disease*, *49*(3), 617–631. https://doi.org/10.3233/JAD-150692

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification* (2nd ed.). John Wiley & Sons, Inc.

Ellis, K. A., Bush, A. I., Darby, D., De Fazio, D., Foster, J., Hudson, P., Lautenschlager, N. T., Lenzo, N., Martins, R. N., Maruff, P., Masters, C., Milner, A., Pike, K., Rowe, C., Savage, G., Szoeke, C., Taddei, K., Villemagne, V., Woodward, M., … AIBL Research Group. (2009). The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *International Psychogeriatrics*, *21*(4), 672–687. https://doi.org/10.1017/S1041610209009405

Escudero, J., Ifeachor, E., & Zajicek, J. P. (2012). Bioprofile analysis: a new approach for the analysis of biomedical data in Alzheimer's disease. *Journal of Alzheimer's Disease*, *32*(4), 997–1010. https://doi.org/10.3233/JAD-2012-121024

Eysenbach, G. (2005). The law of attrition. *Journal of Medical Internet Research*, *7*(1), e11. https://doi.org/10.2196/JMIR.7.1.E11

Fabrizio, C., Termine, A., Caltagirone, C., & Sancesario, G. (2021). Artificial Intelligence for Alzheimer's Disease: Promise or Challenge? *Diagnostics*, *11*(8), 1473.

https://doi.org/10.3390/DIAGNOSTICS11081473

Fotuhi, M., Do, D., & Jack, C. (2012). Modifiable factors that alter the size of the hippocampus with ageing. *Nature Reviews Neurology*, *8*(4), 189–202. https://doi.org/10.1038/nrneurol.2012.27

Fowler, C., Rainey-Smith, S. R., Bird, S., Bomke, J., Bourgeat, P., Brown, B. M., Burnham, S. C., Bush, A. I., Chadunow, C., Collins, S., Doecke, J., Doré, V., Ellis, K. A., Evered, L., Fazlollahi, A., Fripp, J., Gardener, S. L., Gibson, S., Grenfell, R., … Ames, D. (2021). Fifteen Years of the Australian Imaging, Biomarkers and Lifestyle (AIBL) Study: Progress and Observations from 2,359 Older Adults Spanning the Spectrum from Cognitive Normality to Alzheimer's Disease. *Journal of Alzheimer's Disease Reports*, *5*(1), 443–468. https://doi.org/10.3233/ADR-210005

Fox, N. C., & Schott, J. M. (2004). Imaging cerebral atrophy: normal ageing to Alzheimer's disease. *The Lancet*, *363*(9406), 392–394. https://doi.org/10.1016/S0140-6736(04)15441-X

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P., & Thompson, P. M. (2010). The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology*, *6*, 67–77. https://doi.org/10.1038/nrneurol.2009.215

Galimberti, D., & Scarpini, E. (2011). Disease-modifying treatments for Alzheimer's disease. *Therapeutic Advances in Neurological Disorders*, *4*(4), 203–216. https://doi.org/10.1177/1756285611404470

Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R. C., Ritchie, K., Broich, K., Belleville, S., Brodaty, H., Bennett, D., Chertkow, H., Cummings, J. L., de Leon, M., Feldman, H., Ganguli, M., Hampel, H., Scheltens, P., Tierney, M. C., Whitehouse, P., Winblad, B., & International Psychogeriatric Association Expert Conference on mild cognitive impairment. (2006). Mild cognitive impairment. *The Lancet*, *367*(9518), 1262–1270. https://doi.org/10.1016/S0140-6736(06)68542-5

Gencoglu, O., van Gils, M., Guldogan, E., Morikawa, C., Süzen, M., Gruber, M., Leinonen, J., & Huttunen, H. (2019). *HARK Side of Deep Learning -- From Grad Student Descent to Automated Machine Learning*. https://doi.org/10.48550/arxiv.1904.07633

Golriz Khatami, S., Robinson, C., Birkenbihl, C., Domingo-Fernández, D., Hoyt, C. T., & Hofmann-Apitius, M. (2020). Challenges of Integrative Disease Modeling in Alzheimer's Disease. *Frontiers in Molecular Biosciences*, *6*, 158. https://doi.org/10.3389/FMOLB.2019.00158

Gómez-Ramírez, J., Ávila-Villanueva, M., & Fernández-Blázquez, M. Á. (2020). Selecting the most important self-assessed features for predicting conversion to mild cognitive impairment with random forest and permutation-based methods. *Scientific Reports*, *10*, 20630. https://doi.org/10.1038/S41598-020-77296-4

Gomperts, S. N. (2016). Lewy Body Dementias: Dementia With Lewy Bodies and Parkinson Disease Dementia. *Continuum (Minneapolis, Minn.)*, *22*(2 Dementia), 435–463. https://doi.org/10.1212/CON.0000000000000309

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Graham, S. A., Lee, E. E., Jeste, D. V., Van Patten, R., Twamley, E. W., Nebeker, C., Yamada, Y., Kim, H. C., & Depp, C. A. (2020). Artificial intelligence approaches to predicting and detecting cognitive decline in older adults: A conceptual review. *Psychiatry Research*, *284*, 112732. https://doi.org/10.1016/J.PSYCHRES.2019.112732

Guan, H., Liu, Y., Xiao, S., Yue, L., & Liu, M. (2021). Cost-Sensitive Meta-learning for Progress Prediction of Subjective Cognitive Decline with Brain Structural MRI. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. MICCAI 2021. Lecture Notes in Computer Science*, *12905*, 248–258. https://doi.org/10.1007/978-3-030-87240-3_24

Gustaw-Rothenberg, K., Lerner, A., Bonda, D. J., Lee, H., Zhu, X., Perry, G., & Smith, M. A. (2010). Biomarkers in Alzheimer's disease: past, present and future. *Biomarkers in Medicine*, *4*(1), 15–26. https://doi.org/10.2217/bmm.09.86

Hampel, H., Bürger, K., Teipel, S. J., Bokde, A. L. W., Zetterberg, H., & Blennow, K. (2008). Core candidate neurochemical and imaging biomarkers of Alzheimer's disease. *Alzheimer's & Dementia*, *4*(1), 38–48. https://doi.org/10.1016/J.JALZ.2007.08.006

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed., pp. 587–604). Springer. https://doi.org/10.1007/978-0-387-84858-7

Henneman, W. J. P., Sluimer, J. D., Cordonnier, C., Baak, M. M. E., Scheltens, P., Barkhof, F., & van der Flier, W. M. (2009). MRI biomarkers of vascular damage and atrophy predicting mortality in a memory clinic population. *Stroke*, *40*(2), 492–498. https://doi.org/10.1161/STROKEAHA.108.516286

Henneman, W. J. P., Vrenken, H., Barnes, J., Sluimer, I. C., Verwey, N. A., Blankenstein, M. A., Klein, M., Fox, N. C., Scheltens, P., Barkhof, F., & van der Flier, W. M. (2009). Baseline CSF p-tau levels independently predict progression of hippocampal atrophy in Alzheimer disease. *Neurology*, *73*(12), 935–940. https://doi.org/10.1212/WNL.0b013e3181b879ac

Henriksen, K., O'Bryant, S. E., Hampel, H., Trojanowski, J. Q., Montine, T. J., Jeromin, A., Blennow, K., Lönneborg, A., Wyss-Coray, T., Soares, H., Bazenet, C., Sjögren, M., Hu, W., Lovestone, S., Karsdal, M. A., & Weiner, M. W. (2014). The future of blood-based biomarkers for Alzheimer's disease. *Alzheimer's & Dementia*, *10*(1), 115–131. https://doi.org/10.1016/J.JALZ.2013.01.013

Hippius, H., & Neundörfer, G. (2003). The discovery of Alzheimer's disease. *Dialogues in Clinical Neuroscience*, *5*(1), 101–108. https://doi.org/10.31887/DCNS.2003.5.1/hhippius

Huang, L., Jin, Y., Gao, Y., Thung, K.-H., Shen, D., & the Alzheimer's Disease Neuroimaging Initiative. (2016). Longitudinal clinical score prediction in Alzheimer's disease with soft-split sparse regression based random forest. *Neurobiology of Aging*, *46*, 180–191. https://doi.org/10.1016/j.neurobiolaging.2016.07.005

Hui, J. S., Wilson, R. S., Bennett, D. A., Bienias, J. L., Gilley, D. W., & Evans, D. A. (2003). Rate of cognitive decline and mortality in Alzheimer's disease. *Neurology*, *61*(10), 1356–1361. https://doi.org/10.1212/01.WNL.0000094327.68399.59

Iadecola, C. (2013). The Pathobiology of Vascular Dementia. *Neuron*, *80*(4), 844–866.

https://doi.org/10.1016/j.neuron.2013.10.008

Impedovo, D., & Pirlo, G. (2018). Dynamic Handwriting Analysis for the Assessment of Neurodegenerative Diseases: A Pattern Recognition Perspective. *IEEE Reviews in Biomedical Engineering*, *12*, 209–220. https://doi.org/10.1109/RBME.2018.2840679

Ito, K., Corrigan, B., Zhao, Q., French, J., Miller, R., Soares, H., Katz, E., Nicholas, T., Billing, B., Anziano, R., & Fullerton, T. (2011). Disease progression model for cognitive deterioration from Alzheimer's Disease Neuroimaging Initiative database. *Alzheimer's & Dementia*, *7*(2), 151–160. https://doi.org/10.1016/J.JALZ.2010.03.018

Jack, C. R., Albert, M. S., Knopman, D. S., McKhann, G. M., Sperling, R. A., Carrillo, M. C., Thies, B., Phelps, C. H., & Phelps, C. H. (2011). Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 257–262. https://doi.org/10.1016/j.jalz.2011.03.004

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J. L., Montine, T., Phelps, C., Rankin, K. P., Rowe, C. C., Scheltens, P., Siemers, E., Snyder, H. M., … Silverberg, N. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, *14*(4), 535–562. https://doi.org/10.1016/j.jalz.2018.02.018

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Feldman, H. H., Frisoni, G. B., Hampel, H., Jagust, W. J., Johnson, K. A., Knopman, D. S., Petersen, R. C., Scheltens, P., Sperling, R. A., & Dubois, B. (2016). A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology*, *87*(5), 539–547. https://doi.org/10.1212/WNL.0000000000002923

Jack, C. R., Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., Shaw, L. M., Vemuri, P., Wiste, H. J., Weigand, S. D., Lesnick, T. G., Pankratz, V. S., Donohue, M. C., & Trojanowski, J. Q. (2013). Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *The Lancet Neurology*, *12*(2), 207–216. https://doi.org/10.1016/S1474-4422(12)70291-0

Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., Petersen, R. C., & Trojanowski, J. Q. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology*, *9*(1), 119–128. https://doi.org/10.1016/S1474-4422(09)70299-6

Jakob, R., Harperink, S., Rudolf, A. M., Fleisch, E., Haug, S., Mair, J. L., Salamanca-Sanabria, A., & Kowatsch, T. (2022). Factors Influencing Adherence to mHealth Apps for Prevention or Management of Noncommunicable Diseases: Systematic Review. *Journal of Medical Internet Research*, *24*(5), e35371. https://doi.org/10.2196/35371

Jellinger, K. A., & Korczyn, A. D. (2018). Are dementia with Lewy bodies and Parkinson's disease dementia the same disease? *BMC Medicine*, *16*(1), 34. https://doi.org/10.1186/s12916-018-1016-8

Jessen, F., Amariglio, R. E., van Boxtel, M., Breteler, M., Ceccaldi, M., Chételat, G., Dubois, B., Dufouil, C., Ellis, K. A., van Der Flier, W. M., Glodzik, L., van Harten, A. C., de Leon, M. J., McHugh, P., Mielke,

M. M., Molinuevo, J. L., Mosconi, L., Osorio, R. S., Perrotin, A., … Wagner, M. (2014). A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimer's & Dementia*, *10*(6), 844–852. https://doi.org/10.1016/j.jalz.2014.01.001

Jiménez-Huete, A., Riva, E., Toledano, R., Campo, P., Esteban, J., De Barrio, A., & Franch, O. (2014). Differential diagnosis of degenerative dementias using basic neuropsychological tests: multivariable logistic regression analysis of 301 patients. *American Journal of Alzheimer's Disease and Other Dementias*, *29*(8), 723–731. https://doi.org/10.1177/1533317514534954

Jo, T., Nho, K., & Saykin, A. J. (2019). Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. *Frontiers in Aging Neuroscience*, *11*, 220. https://doi.org/10.3389/FNAGI.2019.00220/BIBTEX

Kalaria, R. N. (2018). The pathology and pathophysiology of vascular dementia. *Neuropharmacology*, *134*, 226–239. https://doi.org/10.1016/J.NEUROPHARM.2017.12.030

Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, *64*(5), 402. https://doi.org/10.4097/KJAE.2013.64.5.402

Karantzoulis, S., & Galvin, J. E. (2014). Distinguishing Alzheimer's disease from other major forms of dementia. *Expert Review of Neurotherapeutics*, *11*(11), 1579–1591. https://doi.org/10.1586/ern.11.155

Kivipelto, M., Mangialasche, F., & Ngandu, T. (2018). Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease. *Nature Reviews. Neurology*, *14*(11), 653–666. https://doi.org/10.1038/S41582-018-0070-3

Koenig, A. M., Nobuhara, C. K., Williams, V. J., & Arnold, S. E. (2018). Biomarkers in Alzheimer's, Frontotemporal, Lewy Body, and Vascular Dementias. *Focus*, *16*, 164–172. https://doi.org/10.1176/appi.focus.20170048

Koikkalainen, J., Lötjönen, J., Thurfjell, L., Rueckert, D., Waldemar, G., Soininen, H., & the Alzheimer's Disease Neuroimaging Initiative. (2011). Multi-template tensor-based morphometry: Application to analysis of Alzheimer's disease. *NeuroImage*, *56*(3), 1134–1144. https://doi.org/10.1016/j.neuroimage.2011.03.029

Koikkalainen, J., Rhodius-Meester, H., Tolonen, A., Barkhof, F., Tijms, B., Lemstra, A. W., Tong, T., Guerrero, R., Schuh, A., Ledig, C., Rueckert, D., Soininen, H., Remes, A. M., Waldemar, G., Hasselbalch, S., Mecocci, P., van der Flier, W., & Lötjönen, J. (2016). Differential diagnosis of neurodegenerative diseases using structural MRI data. *NeuroImage: Clinical*, *11*, 435–449. https://doi.org/10.1016/J.NICL.2016.02.019

Korczyn, A. D., Vakhapova, V., & Grinberg, L. T. (2012). Vascular dementia. *Journal of the Neurological Sciences*, *322*(1–2), 2–10. https://doi.org/10.1016/J.JNS.2012.03.027

Kornhuber, J., Schmidtke, K., Frolich, L., Perneczky, R., Wolf, S., Hampel, H., Jessen, F., Heuser, I., Peters, O., Weih, M., Jahn, H., Luckhaus, C., Hüll, M., Gertz, H.-J., Schröder, J., Pantel, J., Rienhoff, O., Seuchter, S. A., Rüther, E., … Wiltfang, J. (2009). Early and differential diagnosis of dementia and mild cognitive impairment: design and cohort baseline characteristics of the German Dementia Competence Network. *Dementia and Geriatric Cognitive Disorders*, *27*(5), 404–417.

https://doi.org/10.1159/000210388

Kühnel, L., Berger, A. K., Markussen, B., & Raket, L. L. (2021). Simultaneous modeling of Alzheimer's disease progression via multiple cognitive scales. *Statistics in Medicine*, *40*(14), 3251–3266. https://doi.org/10.1002/SIM.8932

Kumar, S., Oh, I., Schindler, S., Lai, A. M., Payne, P. R. O., & Gupta, A. (2021). Machine learning for modeling the progression of Alzheimer disease dementia using clinical data: a systematic literature review. *JAMIA Open*, *4*(3), ooab052. https://doi.org/10.1093/JAMIAOPEN/OOAB052

Lähteenmäki, J., Vuorinen, A.-L., Pajula, J., Harno, K., Lehto, M., Niemi, M., & van Gils, M. (2022). Integrating data from multiple Finnish biobanks and national health-care registers for retrospective studies: Practical experiences: *Scandinavian Journal of Public Health*, *50*(4), 482–489. https://doi.org/10.1177/14034948211004421

Lähteenmäki, J., Vuorinen, A. L., Pajula, J., Harno, K., Lehto, M., Niemi, M., & van Gils, M. (2021). Pharmacogenetics of Bleeding and Thromboembolic Events in Direct Oral Anticoagulant Users. *Clinical Pharmacology and Therapeutics*, *110*(3), 768–776. https://doi.org/10.1002/CPT.2316

Lehmann, S., Delaby, C., Touchon, J., Hirtz, C., & Gabelle, A. (2013). Biomarkers of Alzheimer's disease: The present and the future. *Revue Neurologique*, *169*(10), 719–723. https://doi.org/10.1016/j.neurol.2013.07.012

Lewczuk, P., Riederer, P., O'Bryant, S. E., Verbeek, M. M., Dubois, B., Visser, P. J., Jellinger, K. A., Engelborghs, S., Ramirez, A., Parnetti, L., Jack, C. R., Teunissen, C. E., Hampel, H., Lleó, A., Jessen, F., Glodzik, L., de Leon, M. J., Fagan, A. M., Molinuevo, J. L., … on Behalf of the Members of the WFSBP Task Force Working on this Topic: Peter Riederer, Carla Gallo, Dimitrios Kapogiannis, Andrea Lopez Mato, Florence Thibaut. (2018). Cerebrospinal fluid and blood biomarkers for neurodegenerative dementias: An update of the Consensus of the Task Force on Biological Markers in Psychiatry of the World Federation of Societies of Biological Psychiatry. *The World Journal of Biological Psychiatry*, *19*(4), 244–328. https://doi.org/10.1080/15622975.2017.1375556

Li, A., Yue, L., Xiao, S., & Liu, M. (2022). Cognitive Function Assessment and Prediction for Subjective Cognitive Decline and Mild Cognitive Impairment. *Brain Imaging and Behavior*, *16*(2), 645–658. https://doi.org/10.1007/S11682-021-00545-1/TABLES/13

Lindgren, H. (2008). Decision support System Supporting Clinical Reasoning Process – an Evaluation Study in Dementia Care. *Studies in Health Technology and Informatics*, *136*, 315–320.

Lindgren, H. (2011). Towards personalized decision support in the dementia domain based on clinical practice guidelines. *User Modeling and User-Adapted Interaction*, *21*, 377–406. https://doi.org/10.1007/S11257-010-9090-4

Lindgren, H., Eklund, P., & Eriksson, S. (2002). Clinical Decision Support System in Dementia Care. *Studies in Health Technology and Informatics*, *90*, 568–571. https://doi.org/10.3233/978-1-60750-934-9-568

Liu, Y., Pan, Y., Yang, W., Ning, Z., Yue, L., Liu, M., & Shen, D. (2020). Joint Neuroimage Synthesis and Representation Learning for Conversion Prediction of Subjective Cognitive Decline. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. MICCAI 2020. Lecture Notes in Computer*

Science, *12267*, 583–592. https://doi.org/10.1007/978-3-030-59728-3_57

Liu, Y., Yue, L., Xiao, S., Yang, W., Shen, D., & Liu, M. (2022). Assessing clinical progression from subjective cognitive decline to mild cognitive impairment with incomplete multi-modal neuroimages. *Medical Image Analysis*, *75*, 102266. https://doi.org/10.1016/J.MEDIA.2021.102266

Lopez, O. L., Becker, J. T., Wahed, A. S., Saxton, J., Sweet, R. A., Wolk, D. A., Klunk, W., & DeKosky, S. T. (2009). Long-term effects of the concomitant use of memantine with cholinesterase inhibition in Alzheimer disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *80*(6), 600–607. https://doi.org/10.1136/JNNP.2008.158964

Lötjönen, J., Ledig, C., Koikkalainen, J., Wolz, R., Thurfjell, L., Soininen, H., Ourselin, S., Rueckert, D., & the Alzheimer's Disease Neuroimaging Initiative. (2014). Extended boundary shift integral. *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, 854–857. https://doi.org/10.1109/ISBI.2014.6868005

Lötjönen, J., Wolz, R., Koikkalainen, J. R., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D., & the Alzheimer's Disease Neuroimaging Initiative. (2010). Fast and robust multi-atlas segmentation of brain magnetic resonance images. *NeuroImage*, *49*(3), 2352–2365. https://doi.org/10.1016/j.neuroimage.2009.10.026

Mank, A., van Maurik, I. S., Rijnhart, J. J. M., Bakker, E. D., Bouteloup, V., Le Scouarnec, L., Teunissen, C. E., Barkhof, F., Scheltens, P., Berkhof, J., & van der Flier, W. M. (2022). Development of multivariable prediction models for institutionalization and mortality in the full spectrum of Alzheimer's disease. *Alzheimer's Research and Therapy*, *14*(1), 1–12. https://doi.org/10.1186/s13195-022-01053-0

Mann, D. M. A., & Snowden, J. S. (2017). Frontotemporal lobar degeneration: Pathogenesis, pathology and pathways to phenotype. *Brain Pathology*, *27*(6), 723–736. https://doi.org/10.1111/bpa.12486

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press. https://doi.org/10.1017/CBO9780511809071

Marinescu, R. V., Oxtoby, N. P., Young, A. L., Bron, E. E., Toga, A. W., Weiner, M. W., Barkhof, F., Fox, N. C., Golland, P., Klein, S., & Alexander, D. C. (2019). TADPOLE Challenge: Accurate Alzheimer's Disease Prediction Through Crowdsourced Forecasting of Future Data. In S. H. Rekik, Islem; Adeli, Ehsan; Park (Ed.), *Predictive Intelligence in Medicine: Vol. LNCS 11843* (pp. 1–10). Springer. https://doi.org/10.1007/978-3-030-32281-6_1

Martí-Juan, G., Sanroma-Guell, G., & Piella, G. (2020). A survey on machine and statistical learning for longitudinal analysis of neuroimaging data in Alzheimer's disease. *Computer Methods and Programs in Biomedicine*, *189*, 105348. https://doi.org/10.1016/J.CMPB.2020.105348

MathWorks. (2021). *fitcnb*. https://se.mathworks.com/help/stats/fitcnb.html#budugq6-13

Mattila, J. (2014). *Disease state index and disease state fingerprint: supervised learning applied to clinical decision support in Alzheimer's disease* [Tampere University of Technology]. https://trepo.tuni.fi/handle/10024/115061

Mattila, J., Koikkalainen, J., Virkki, A., Simonsen, A., van Gils, M., Waldemar, G., Soininen, H., Lötjönen,

J., & the Alzheimer's Disease Neuroimaging Initiative. (2011). A disease state fingerprint for evaluation of Alzheimer's disease. *Journal of Alzheimer's Disease*, *27*(1), 163–176. https://doi.org/10.3233/JAD-2011-110365

Mattila, J., Koikkalainen, J., Virkki, A., van Gils, M., & Lötjönen, J. (2012). Design and application of a generic clinical decision support system for multiscale data. *IEEE Transactions on Biomedical Engineering*, *59*(1), 234–240. https://doi.org/10.1109/TBME.2011.2170986

Maurer, K., Volk, S., & Gerbaldo, H. (1997). Auguste D and Alzheimer's disease. *The Lancet*, *349*(9064), 1546–1549. https://doi.org/10.1016/S0140-6736(96)10203-8

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*(7), 939–944. https://doi.org/10.1212/WNL.34.7.939

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C., Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S., & Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 263–269. https://doi.org/10.1016/j.jalz.2011.03.005

Meester, H. F. M. R.-, Koikkalainen, J., Paajanen, T., Mahdiani, S., Barkhof, F., Herukka, S.-K., Hänninen, T., Ngandu, T., Kivipelto, M., Gils, M. van, Hasselbalch, S. G., Mecocci, P., Remes, A., Soininen, H., Scheltens, P., Flier, W. van Der, & Lötjönen, J. (2020). Differential diagnosis of dementia combining web-based cognitive testing and MRI. *Alzheimer's & Dementia*, *16*(S6), e042626. https://doi.org/10.1002/ALZ.042626

Molinuevo, J. L., Rabin, L. A., Amariglio, R., Buckley, R., Dubois, B., Ellis, K. A., Ewers, M., Hampel, H., Klöppel, S., Rami, L., Reisberg, B., Saykin, A. J., Sikkes, S., Smart, C. M., Snitz, B. E., Sperling, R., van der Flier, W. M., Wagner, M., & Jessen, F. (2017). Implementation of subjective cognitive decline criteria in research studies. *Alzheimer's & Dementia*, *13*(3), 296–311. https://doi.org/10.1016/J.JALZ.2016.09.012

Mouiha, A., Duchesne, S., & Initiative, the A. D. N. (2012). Toward a Dynamic Biomarker Model in Alzheimer's Disease. *Journal of Alzheimer's Disease*, *30*(1), 91–100. https://doi.org/10.3233/JAD-2012-111367

Muñoz-Ruiz, M. Á., Hall, A., Mattila, J., Koikkalainen, J., Herukka, S. K., Husso, M., Hänninen, T., Vanninen, R., Liu, Y., Hallikainen, M., Lötjönen, J., Remes, A. M., Alafuzoff, I., Soininen, H., & Hartikainen, P. (2016). Using the Disease State Fingerprint Tool for Differential Diagnosis of Frontotemporal Dementia and Alzheimer's Disease. *Dementia and Geriatric Cognitive Disorders Extra*, *6*(2), 313–329. https://doi.org/10.1159/000447122

Muurling, M., Rhodius-Meester, H. F. M., Pärkkä, J., Van Gils, M., Frederiksen, K. S., Bruun, M., Hasselbalch, S. G., Soininen, H., Herukka, S. K., Hallikainen, M., Teunissen, C. E., Visser, P. J., Scheltens, P., Van Der Flier, W. M., Mattila, J., Lötjönen, J., & De Boer, C. (2020). Gait Disturbances

are Associated with Increased Cognitive Impairment and Cerebrospinal Fluid Tau Levels in a Memory Clinic Cohort. *Journal of Alzheimer's Disease*, *76*(3), 1061–1070. https://doi.org/10.3233/JAD-200225

Nägga, K., Wattmo, C., Zhang, Y., Wahlund, L.-O., & Palmqvist, S. (2014). Cerebral inflammation is an underlying mechanism of early death in Alzheimer's disease: A 13-year cause-specific multivariate mortality study. *Alzheimer's Research and Therapy*, *6*(41). https://doi.org/10.1186/alzrt271

Ngandu, T., Lehtisalo, J., Solomon, A., Levälahti, E., Ahtiluoto, S., Antikainen, R., Bäckman, L., Hänninen, T., Jula, A., Laatikainen, T., Lindström, J., Mangialasche, F., Paajanen, T., Pajala, S., Peltonen, M., Rauramaa, R., Stigsdotter-Neely, A., Strandberg, T., Tuomilehto, J., … Kivipelto, M. (2015). A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): A randomised controlled trial. *The Lancet*, *385*(9984), 2255–2263. https://doi.org/10.1016/S0140-6736(15)60461-5

Nordström, P., Religa, D., Wimo, A., Winblad, B., & Eriksdotter, M. (2013). The use of cholinesterase inhibitors and the risk of myocardial infarction and death: a nationwide cohort study in subjects with Alzheimer's disease. *European Heart Journal*, *34*(33), 2585–2591. https://doi.org/10.1093/EURHEARTJ/EHT182

O'Brien, J. T., Erkinjuntti, T., Reisberg, B., Roman, G., Sawada, T., Pantoni, L., Bowler, J. V, Ballard, C., DeCarli, C., Gorelick, P. B., Rockwood, K., Burns, A., Gauthier, S., & DeKosky, S. T. (2003). Vascular cognitive impairment. *Lancet Neurology*, *2*, 89–98. https://doi.org/10.1016/s1474-4422(03)00305-3

O'Brien, J. T., & Thomas, A. (2015). Vascular dementia. *The Lancet*, *386*(10004), 1698–1706. https://doi.org/10.1016/S0140-6736(15)00463-8

Perani, D., Cerami, C., Caminiti, S. P., Santangelo, R., Coppi, E., Ferrari, L., Pinto, P., Passerini, G., Falini, A., Iannaccone, S., Cappa, S. F., Comi, G., Gianolli, L., & Magnani, G. (2016). Cross-validation of biomarkers for the early differential diagnosis and prognosis of dementia in a clinical setting. *European Journal of Nuclear Medicine and Molecular Imaging*, *43*(3), 499–508. https://doi.org/10.1007/S00259-015-3170-Y

Pereira, J. B., Janelidze, S., Stomrud, E., Palmqvist, S., van Westen, D., Dage, J. L., Mattsson-Carlgren, N., & Hansson, O. (2021). Plasma markers predict changes in amyloid, tau, atrophy and cognition in non-demented subjects. *Brain*, *144*(9), 2826–2836. https://doi.org/10.1093/brain/awab163

Petersen, R. C. (2009). Early diagnosis of Alzheimer's disease: is MCI too late? *Current Alzheimer Research*, *6*(4), 324–330. https://doi.org/10.2174/156720509788929237

Petersen, R. C. (2016). Mild Cognitive Impairment. *Continuum (Minneap Minn)*, *22*(2), 404–418. https://doi.org/10.1212/CON.0000000000000313

Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., Jack, C. R., Jagust, W. J., Shaw, L. M., Toga, A. W., Trojanowski, J. Q., & Weiner, M. W. (2010). Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization. *Neurology*, *74*(3), 201. https://doi.org/10.1212/WNL.0B013E3181CB3E25

Prince, M., Wimo, A., Guerchet, M., Ali, G.-C., Wu, Y.-T., Prina, M., & Alzheimer's Disease International. (2015). *World Alzheimer Report 2015: The Global Impact of Dementia: An Analysis of prevalence, incidence, cost and trends*. www.alz.co.uk/worldreport2015corrections

Qiu, S., Joshi, P. S., Miller, M. I., Xue, C., Zhou, X., Karjadi, C., Chang, G. H., Joshi, A. S., Dwyer, B., Zhu, S., Kaku, M., Zhou, Y., Alderazi, Y. J., Swaminathan, A., Kedar, S., Saint-Hilaire, M. H., Auerbach, S. H., Yuan, J., Sartor, E. A., … Kolachalama, V. B. (2020). Development and validation of an interpretable deep learning framework for Alzheimer's disease classification. *Brain*, *143*(6), 1920–1933. https://doi.org/10.1093/BRAIN/AWAA137

Raket, L. L. (2020). Statistical Disease Progression Modeling in Alzheimer Disease. *Frontiers in Big Data*, *3*, 24. https://doi.org/10.3389/fdata.2020.00024

Raskin, J., Cummings, J., Hardy, J., Schuh, K., & Dean, R. (2015). Neurobiology of Alzheimer's Disease: Integrated Molecular, Physiological, Anatomical, Biomarker, and Cognitive Dimensions. *Current Alzheimer Research*, *12*(8), 712–722. https://doi.org/10.2174/1567205012666150701103107

Rathore, S., Habes, M., Iftikhar, M. A., Shacklett, A., & Davatzikos, C. (2017). A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. *NeuroImage*, *155*, 530–548. https://doi.org/10.1016/J.NEUROIMAGE.2017.03.057

Rountree, S. D., Chan, W., Pavlik, V. N., Darby, E. J., & Doody, R. S. (2012). Factors that influence survival in a probable Alzheimer disease cohort. *Alzheimer's Research & Therapy*, *4*(3), 16. https://doi.org/10.1186/ALZRT119

Samtani, M. N., Farnum, M., Lobanov, V., Yang, E., Raghavan, N., Dibernardo, A., & Narayan, V. (2012). An improved model for disease progression in patients from the Alzheimer's disease neuroimaging initiative. *Journal of Clinical Pharmacology*, *52*(5), 629–644. https://doi.org/10.1177/0091270011405497

Serrano-Pozo, A., Frosch, M. P., Masliah, E., & Hyman, B. T. (2011). Neuropathological alterations in Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine*, *1*, a006189. https://doi.org/10.1101/cshperspect.a006189

Shaw, L. M., Korecka, M., Clark, C. M., Lee, V. M.-Y., & Trojanowski, J. Q. (2007). Biomarkers of neurodegeneration for diagnosis and monitoring therapeutics. *Nature Reviews Drug Discovery*, *6*(4), 295–303. https://doi.org/10.1038/nrd2176

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., Lillicrap, T., Simonyan, K., & Hassabis, D. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science*, *362*(6419), 1140–1144. https://doi.org/10.1126/SCIENCE.AAR6404

Sim, I., Gorman, P., Greenes, R. A., Haynes, R. B., Kaplan, B., Lehmann, H., & Tang, P. C. (2001). Clinical decision support systems for the practice of evidence-based medicine. *Journal of the American Medical Informatics Association*, *8*(6), 527–534. https://doi.org/10.1136/jamia.2001.0080527

Simonsen, A. H., Mattila, J., Hejl, A. M., Frederiksen, K. S., Herukka, S. K., Hallikainen, M., van Gils, M., Lötjönen, J., Soininen, H., & Waldemar, G. (2013). Application of the predictad software tool to

predict progression in patients with mild cognitive impairment. *Dementia and Geriatric Cognitive Disorders*, *34*(5–6), 344–350. https://doi.org/10.1159/000345554

Sloane, P. D., Zimmerman, S., Suchindran, C., Reed, P., Wang, L., Boustani, M., & Sudha, S. (2002). The public health impact of Alzheimer's disease, 2000-2050: Potential implication of treatment advances. *Annual Review of Public Health*, *23*, 213–231. https://doi.org/10.1146/annurev.publhealth.23.100901.140525

Soininen, H., Mattila, J., Koikkalainen, J., van Gils, M., Hviid Simonsen, A., Waldemar, G., Rueckert, D., Thurfjell, L., & Lötöjnen, J. (2012). Software tool for improved prediction of Alzheimer's disease. *Neurodegenerative Diseases*, *10*(1–4), 149–152. https://doi.org/10.1159/000332600

Speechly, C. M., Bridges-Webb, C., & Passmore, E. (2008). The pathway to dementia diagnosis. *The Medical Journal of Australia*, *189*(9), 487–489. https://doi.org/10.5694/J.1326-5377.2008.TB02140.X

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., Iwatsubo, T., Jack, C. R., Kaye, J., Montine, T. J., Park, D. C., Reiman, E. M., Rowe, C. C., Siemers, E., Stern, Y., Yaffe, K., Carrillo, M. C., Thies, B., Morrison-Bogorad, M., … Phelps, C. H. (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. *Alzheimer's & Dementia*, *7*(3), 280–292. https://doi.org/10.1016/j.jalz.2011.03.003

Sperling, R., & Johnson, K. (2013). Biomarkers of Alzheimer Disease: Current and Future Applications to Diagnostic Criteria. *Continuum (Minneap Minn)*, *19*(2), 325–338. https://doi.org/10.1212/01.CON.0000429181.60095.99

Stern, Y. (2009). Cognitive reserve. *Neuropsychologia*, *47*(10), 2015–2028. https://doi.org/10.1016/j.neuropsychologia.2009.03.004

Stern, Y. (2012). Cognitive reserve in ageing and Alzheimer's disease. *The Lancet Neurology*, *11*(11), 1006–1012. https://doi.org/10.1016/S1474-4422(12)70191-6

Stricker, N. H., Dodge, H. H., Dowling, N. M., Han, S. D., Erosheva, E. A., Jagust, W. J., & the Alzheimer's Disease Neuroimaging Initiative. (2012). CSF biomarker associations with change in hippocampal volume and precuneus thickness: implications for the Alzheimer's pathological cascade. *Brain Imaging and Behavior*, *6*(4), 599–609. https://doi.org/10.1007/s11682-012-9171-6

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press. https://www.andrew.cmu.edu/course/10-703/textbook/BartoSutton.pdf

Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digital Medicine*, *3*, 17. https://doi.org/10.1038/s41746-020-0221-y

Tan, M. S., Cheah, P. L., Chin, A. V., Looi, L. M., & Chang, S. W. (2021). A review on omics-based biomarkers discovery for Alzheimer's disease from the bioinformatics perspectives: Statistical approach vs machine learning approach. *Computers in Biology and Medicine*, *139*, 104947. https://doi.org/10.1016/J.COMPBIOMED.2021.104947

Tăuţan, A. M., Ionescu, B., & Santarnecchi, E. (2021). Artificial intelligence in neurodegenerative diseases: A review of available tools with a focus on machine learning techniques. *Artificial Intelligence in Medicine*, *117*, 102081. https://doi.org/10.1016/J.ARTMED.2021.102081

Teunissen, C. E., Verberk, I. M. W., Thijssen, E. H., Vermunt, L., Hansson, O., Zetterberg, H., van der Flier, W. M., Mielke, M. M., & del Campo, M. (2022). Blood-based biomarkers for Alzheimer's disease: towards clinical implementation. *The Lancet Neurology*, *21*(1), 66–77. https://doi.org/10.1016/S1474-4422(21)00361-6

Todd, S., Barr, S., Roberts, M., & Passmore, A. P. (2013). Survival in dementia and predictors of mortality: a review. *International Journal of Geriatric Psychiatry*, *28*(11), 1109–1124. https://doi.org/10.1002/GPS.3946

Tohka, J., & van Gils, M. (2021). Evaluation of machine learning algorithms for health and wellness applications: A tutorial. *Computers in Biology and Medicine*, *132*, 104324. https://doi.org/10.1016/j.compbiomed.2021.104324

Tolonen, A., Rhodius-Meester, H. F. M., Bruun, M., Koikkalainen, J., Barkhof, F., Lemstra, A. W., Koene, T., Scheltens, P., Teunissen, C. E., Tong, T., Guerrero, R., Schuh, A., Ledig, C., Baroni, M., Rueckert, D., Soininen, H., Remes, A. M., Waldemar, G., Hasselbalch, S. G., … Lötjönen, J. (2018). Data-driven differential diagnosis of dementia using multiclass disease state index classifier. *Frontiers in Aging Neuroscience*, *10*, 111. https://doi.org/10.3389/fnagi.2018.00111

Tong, T., Ledig, C., Guerrero, R., Schuh, A., Koikkalainen, J., Tolonen, A., Rhodius, H., Barkhof, F., Tijms, B., Lemstra, A. W., Soininen, H., Remes, A. M., Waldemar, G., Hasselbalch, S., Mecocci, P., Baroni, M., Lötjönen, J., Flier, W. van der, & Rueckert, D. (2017). Five-class differential diagnostics of neurodegenerative diseases using random undersampling boosting. *NeuroImage: Clinical*, *15*, 613–624. https://doi.org/10.1016/J.NICL.2017.06.012

Valech, N., Mollica, M. A., Olives, J., Tort, A., Fortea, J., Lleo, A., Belén, S.-S., Molinuevo, J. L., & Rami, L. (2015). Informants' Perception of Subjective Cognitive Decline Helps to Discriminate Preclinical Alzheimer's Disease from Normal Aging. *Journal of Alzheimer's Disease*, *48*(s1), S87–S98. https://doi.org/10.3233/JAD-150117

van de Pol, L. A., van der Flier, W. M., Korf, E. S. C., Fox, N. C., Barkhof, F., & Scheltens, P. (2007). Baseline predictors of rates of hippocampal atrophy in mild cognitive impairment. *Neurology*, *69*(15), 1491–1497. https://doi.org/10.1212/01.wnl.0000277458.26846.96

van de Vorst, I. E., Koek, H. L., de Vries, R., Bots, M. L., Reitsma, J. B., & Vaartjes, I. (2016). Effect of Vascular Risk Factors and Diseases on Mortality in Individuals with Dementia: A Systematic Review and Meta-Analysis. *Journal of the American Geriatrics Society*, *64*(1), 37–46. https://doi.org/10.1111/JGS.13835

van der Flier, W. M., Pijnenburg, Y. A. L., Prins, N., Lemstra, A. W., Bouwman, F. H., Teunissen, C. E., van Berckel, B. N. M., Stam, C. J., Barkhof, F., Visser, P. J., van Egmond, E., & Scheltens, P. (2014). Optimizing Patient Care and Research: The Amsterdam Dementia Cohort. *Journal of Alzheimer's Disease*, *41*, 313–327. https://doi.org/10.3233/JAD-132306

Vuoksimaa, E., McEvoy, L. K., Holland, D., Franz, C. E., Kremen, W. S., & the Alzheimer's Disease Neuroimaging Initiative. (2020). Modifying the minimum criteria for diagnosing amnestic MCI to improve prediction of brain atrophy and progression to Alzheimer's disease. *Brain Imaging and Behavior*, *14*, 787–796. https://doi.org/10.1007/S11682-018-0019-6

Vuorinen, A. L., Lehto, M., Niemi, M., Harno, K., Pajula, J., van Gils, M., & Lähteenmäki, J. (2021). Pharmacogenetics of Anticoagulation and Clinical Events in Warfarin-Treated Patients: A Register-Based Cohort Study with Biobank Data and National Health Registries in Finland. *Clinical Epidemiology*, *13*, 183. https://doi.org/10.2147/CLEP.S289031

Walker, Z., Possin, K. L., Boeve, B. F., & Aarsland, D. (2015). Lewy body dementias. *The Lancet*, *386*(10004), 1683–1697. https://doi.org/10.1016/S0140-6736(15)00462-6

Wallin, Å. K., Blennow, K., Zetterberg, H., Londos, E., Minthon, L., & Hansson, O. (2010). CSF biomarkers predict a more malignant outcome in Alzheimer disease. *Neurology*, *74*(19), 1531–1537. https://doi.org/10.1212/WNL.0B013E3181DD4DD8

Wang, P., Li, Y., & Reddy, C. K. (2019). Machine Learning for Survival Analysis. *ACM Computing Surveys (CSUR)*, *51*(6). https://doi.org/10.1145/3214306

Wattmo, C., Londos, E., & Minthon, L. (2015). Longitudinal Associations between Survival in Alzheimer's Disease and Cholinesterase Inhibitor Use, Progression, and Community-Based Services. *Dementia and Geriatric Cognitive Disorders*, *40*(5–6), 297–310. https://doi.org/10.1159/000437050

Webb, A. R., & Copsey, K. D. (2011). *Statistical Pattern Recognition* (3rd ed.). John Wiley & Sons, Ltd.

Weiner, M. W., Aisen, P. S., Jack, C. R., Jagust, W. J., Trojanowski, J. Q., Shaw, L., Saykin, A. J., Morris, J. C., Cairns, N., Beckett, L. a, Toga, A., Green, R., Walter, S., Soares, H., Snyder, P., Siemers, E., Potter, W., Cole, P. E., & Schmidt, M. (2010). The Alzheimer's disease neuroimaging initiative: progress report and future plans. *Alzheimer's & Dementia*, *6*, 202-211.e7. https://doi.org/10.1016/j.jalz.2010.03.007

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Cedarbaum, J., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., Luthman, J., Morris, J. C., Petersen, R. C., Saykin, A. J., Shaw, L., Shen, L., Schwarz, A., Toga, A. W., & Trojanowski, J. Q. (2015). 2014 Update of the Alzheimer's Disease Neuroimaging Initiative: A review of papers published since its inception. *Alzheimer's & Dementia*, *11*, e1–e120. https://doi.org/10.1016/J.JALZ.2014.11.001

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., Morris, J. C., Petersen, R. C., Saykin, A. J., Shaw, L. M., Toga, A. W., & Trojanowski, J. Q. (2017). Recent publications from the Alzheimer's Disease Neuroimaging Initiative: Reviewing progress toward improved AD clinical trials. *Alzheimer's & Dementia*, *13*, e1–e85. https://doi.org/10.1016/J.JALZ.2016.11.007

Wilson, R. S., Li, Y., Aggarwal, N. T., McCann, J. J., Gilley, D. W., Bienias, J. L., Barnes, L. L., & Evans, D. A. (2006). Cognitive decline and survival in Alzheimer's disease. *International Journal of Geriatric Psychiatry*, *21*(4), 356–362. https://doi.org/10.1002/GPS.1472

Wolfsgruber, S., Polcher, A., Koppara, A., Kleineidam, L., Frölich, L., Peters, O., Hüll, M., Rüther, E.,

Wiltfang, J., Maier, W., Kornhuber, J., Lewczuk, P., Jessen, F., & Wagner, M. (2017). Cerebrospinal Fluid Biomarkers and Clinical Progression in Patients with Subjective Cognitive Decline and Mild Cognitive Impairment. *Journal of Alzheimer's Disease*, *58*(3), 939–950. https://doi.org/10.3233/JAD-161252

Xu, H., Garcia-Ptacek, S., Jönsson, L., Wimo, A., Nordström, P., & Eriksdotter, M. (2021). Long-term Effects of Cholinesterase Inhibitors on Cognitive Decline and Mortality. *Neurology*, *96*(17), e2220–e2230. https://doi.org/10.1212/WNL.0000000000011832

Yue, L., Hu, D., Zhang, H., Wen, J., Wu, Y., Li, W., Sun, L., Li, X., Wang, J., Li, G., Wang, T., Shen, D., & Xiao, S. (2021). Prediction of 7-year's conversion from subjective cognitive decline to mild cognitive impairment. *Human Brain Mapping*, *42*(1), 192–203. https://doi.org/10.1002/HBM.25216

Zhang, D., Shen, D., & the Alzheimer's Disease Neuroimaging Initiative. (2012a). Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *NeuroImage*, *59*(2), 895–907. https://doi.org/10.1016/j.neuroimage.2011.09.069

Zhang, D., Shen, D., & the Alzheimer's Disease Neuroimaging Initiative. (2012b). Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. *PLoS ONE*, *7*(3), e33182. https://doi.org/10.1371/journal.pone.0033182

Zhang, D., Wang, Y., Zhou, L., Yuan, H., & Shen, D. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *NeuroImage*, *55*(3), 856–867. https://doi.org/10.1016/j.neuroimage.2011.01.008

Zhang, R., Simon, G., & Yu, F. (2017). Advancing Alzheimer's research: A review of big data promises. *International Journal of Medical Informatics*, *106*, 48–56. https://doi.org/10.1016/J.IJMEDINF.2017.07.002

Zhou, J., Liu, J., Narayan, V. A., Ye, J., & the Alzheimer's Disease Neuroimaging Initiative. (2013). Modeling disease progression via multi-task learning. *NeuroImage*, *78*, 233–248. https://doi.org/10.1016/j.neuroimage.2013.03.073

Zhu, F., Panwar, B., Dodge, H. H., Li, H., Hampstead, B. M., Albin, R. L., Paulson, H. L., & Guan, Y. (2016). COMPASS: a computational model to predict changes in MMSE scores 24-months after initial assessment of Alzheimer's disease. *Scientific Reports*, *6*, 34567. https://doi.org/10.1038/srep34567

Zissimopoulos, J., Crimmins, E., & St.clair, P. (2014). The Value of Delaying Alzheimer's Disease Onset. *Forum for Health Economics & Policy*, *18*(1), 25–39. https://doi.org/10.1515/FHEP-2014-0013

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

PUBLICATION I

Computer assisted prediction of clinical progression in the earliest stages of AD

Rhodius-Meester H. F. M, Liedes H., Koikkalainen J., Wolfsgruber S., Coll-Padros N., Kornhuber J., Peters O., Jessen F., Kleineidam L., Molinuevo J. L., Rami L., Teunissen C. E., Barkhof F., Sikkes S. A. M., Wesselman L. M. P., Slot R. E. R., Verfaillie S. C. J., Scheltens P., Tijms B. M., Lötjönen J. & van der Flier W. M.

Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring, 2018, 10, 726-736.
https://doi.org/10.1016/j.dadm.2018.09.001

ELSEVIER

Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 10 (2018) 726-736

Alzheimer's & Dementia

Cognitive & Behavioral Assessment

Computer-assisted prediction of clinical progression in the earliest stages of AD

Hanneke F. M. Rhodius-Meester[a,*], Hilkka Liedes[b], Juha Koikkalainen[b,c], Steffen Wolfsgruber[d], Nina Coll-Padros[e], Johannes Kornhuber[f], Oliver Peters[g], Frank Jessen[h], Luca Kleineidam[d], José Luis Molinuevo[e,i], Lorena Rami[e], Charlotte E. Teunissen[j], Frederik Barkhof[k], Sietske A. M. Sikkes[a], Linda M. P. Wesselman[a], Rosalinde E. R. Slot[a], Sander C. J. Verfaillie[a], Philip Scheltens[a], Betty M. Tijms[a], Jyrki Lötjönen[c], Wiesje M. van der Flier[a,l]

[a]*Alzheimer Center, Department of Neurology, VU University Medical Centre, Amsterdam Neuroscience, Amsterdam, the Netherlands*
[b]*VTT Technical Research Centre of Finland Ltd., Tampere, Finland*
[c]*Combinostics Ltd., Tampere, Finland*
[d]*Department for Neurodegenerative Diseases and Geriatric Psychiatry, University of Bonn, Germany, and German Center for Neurodegenerative Diseases, Bonn, Germany*
[e]*Alzheimer's Disease and Other Cognitive Disorders Unit, Hospital Clínic, IDIBAPS, Barcelona, Spain*
[f]*Department of Psychiatry and Psychotherapy, Universitätsklinikum Erlangen and Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany*
[g]*Department of Psychiatry, Charité Berlin, Campus Benjamin Franklin, Berlin, Germany*
[h]*Department of Psychiatry, University of Cologne, Cologne, Germany*
[i]*Barcelona Beta Brain Research Center, Pasqual Maragall Foundation, Barcelona, Spain*
[j]*Neurochemistry Lab and Biobank, Department of Clinical Chemistry, VU University Medical Centre, Amsterdam Neuroscience, Amsterdam, the Netherlands*
[k]*Department of Radiology and Nuclear Medicine, VU University Medical Centre, Amsterdam Neuroscience, Amsterdam, the Netherlands and Institutes of Neurology and Healthcare Engineering, UCL, London, United Kingdom*
[l]*Department of Epidemiology and Biostatistics, VU University Medical Centre, Amsterdam Neuroscience, Amsterdam, the Netherlands*

Abstract

Introduction: Individuals with subjective cognitive decline (SCD) are at increased risk for clinical progression. We studied how combining different diagnostic tests can help to identify individuals who are likely to show clinical progression.
Methods: We included 674 patients with SCD (46% female, 64 ± 9 years, Mini–Mental State Examination 28 ± 2) from three memory clinic cohorts. A multivariate model based on the Disease State Index classifier incorporated the available baseline tests to predict progression to MCI or dementia over time. We developed and internally validated the model in one cohort and externally validated it in the other cohorts.
Results: After 2.9 ± 2.0 years, 151(22%) patients showed clinical progression. Overall performance of the classifier when combining cognitive tests, magnetic resonance imagining, and cerebrospinal fluid showed a balanced accuracy of 74.0 ± 5.5, with high negative predictive value (93.3 ± 2.8).
Discussion: We found that a combination of diagnostic tests helps to identify individuals at risk of progression. The classifier had particularly good accuracy in identifying patients who remained stable.

Keywords: Alzheimer's disease; Prognosis; Diagnostic test assessment; Clinical decision support system; Subjective cognitive decline

1. Background

In the setting of a memory clinic, patients with subjective cognitive decline (SCD) are highly relevant [1]. Most of

The authors have declared that no conflict of interest exists.
*Corresponding author. Tel.: +31204440816; Fax: +31204448529.
E-mail address: h.rhodius@vumc.nl

https://doi.org/10.1016/j.dadm.2018.09.001

them are "worried well", yet a small proportion of these patients is likely to suffer from preclinical Alzheimer's disease (AD) [2,3]. For both the patient and the clinician, it is important to know who will progress to mild cognitive impairment (MCI) or dementia and who will remain stable [2,4,5].

At this point, cerebrospinal fluid (CSF) and magnetic resonance imagining (MRI) markers, and to a lesser extent cognitive tests, are associated with decline in SCD [3,6–13]. These findings have been translated into the "SCD plus"—criteria that have been developed to identify individuals who are more likely to harbor preclinical AD [2,14]. Translation to clinical practice is hampered because a set of recommendations for what the diagnostic workup and follow-up for patients with SCD should look like is currently lacking [15,16].

Clinical decision support systems based on modern machine-learning technologies are being developed to support clinicians to integrate multiple determinants in daily practice [17]. We have previously developed the Disease State Index (DSI) classifier, which is a technology that integrates patient data from multiple modalities to support the clinician in decision-making [18]. In previous studies, we showed that the DSI can distinguish different types of dementia and discriminate between stable and progressive MCI patients [18–21].

In this study, we aimed to investigate and validate in independent cohorts the prognostic ability of the DSI classifier to identify patients with SCD at risk for progression, by combining and visualizing all available data on baseline characteristics, neuropsychology, CSF biomarkers, and automated MRI features.

2. Methods

2.1. Patients

We included 674 patients with SCD with baseline neuropsychology available and a minimal follow-up of 1 year, from three different memory clinic-based cohorts: 354 from the Amsterdam Dementia Cohort (ADC) from the VU Medical Center [22–24], 51 from Barcelona [25], and 269 from the German Dementia Competence Network (DCN), consisting of nine memory clinics [26,27]. We used the ADC cohort to develop and internally validate our model and the pooled data of the Barcelona and DCN cohorts to externally validate our model. The study was approved by the local medical ethical committees. All patients provided written informed consent for their clinical data to be used for research purposes.

2.2. Clinical assessment

All patients went to the memory clinics seeking medical help. At baseline, they received a standardized and multidisciplinary work-up, including medical history and neuropsychological examination. CSF and MRI were performed in a subset of patients. In multidisciplinary consensus meetings, patients were labeled as having SCD when the cognitive complaints could not be confirmed by cognitive testing using a neuropsychological battery and criteria for MCI, dementia, or other neurologic or psychiatric disorder known to cause cognitive complaints were not met.

Annual follow-up took place by routine clinical visits, in which medical and neuropsychological examinations were repeated. As outcome measure, we defined clinical progression as conversion to MCI, AD, or another type of dementia as diagnosed at follow-up. Time to follow-up was defined as time in years from baseline SCD diagnosis to progression or, if stable, time to most recent follow-up date. In the ADC and Barcelona cohort, MCI was diagnosed using Petersen's criteria; in addition, all patients fulfilled the core clinical criteria of the NIA-AA for MCI [28,29]. In DCN, MCI patients met the Jak and Bondi criteria [30]. Patients were diagnosed with probable AD using the criteria of the NINCDS-ADRDA in all centers; all patients also met the core clinical criteria of the NIA-AA for AD dementia [31,32].

2.3. Neuropsychological tests

Cognitive functions were assessed with a standardized test battery, and we selected those tests that overlapped between the three centers. We used the Mini–Mental State Examination for global cognitive functioning [33]. For measuring executive functioning, we used Trail Making Test A (TMT-A) and Test B (TMT-B), and also for measuring language, category fluency (animals) [34,35]. For episodic memory, we included the tests that resembled each other most. In ADC, the Rey Auditory Verbal Learning Task (RAVLT) immediate and delayed recall were included [36]. In the Barcelona cohort, the Free and Cued Selective Reminding Test (FSCRT) immediate and delayed total recall were used [37]. In DCN, the Consortium to Establish a Registry for Alzheimer's Disease word list immediate and delayed recall were used [38]. To pool the different memory tests, we standardized RAVLT, FSCRT, and Consortium to Establish a Registry for Alzheimer's Disease per center to z-scores using group mean (details on distribution can be found in Supplementary Fig. A1). Missing data varied per test, details can be found in Table 1.

2.4. MRI

In ADC, patients were scanned routinely on a 1.0 T (n = 183), 1.5 T (n = 26), or 3.0 T (n = 123) MRI scanners. Images were acquired on a 3.0 T scanner in Barcelona (n = 49) and on 1.5 T scanners in DCN (n = 93). A set of computed MRI imaging biomarkers were extracted using an image quantification tool (Combinostics Oy, Tampere, Finland, www.cneuro.com/cmri/) [19]. We included four features in the current analysis: hippocampal volume,

Table 1
Baseline characteristics according to outcome at follow-up for the separate centers

	ADC			Barcelona			DCN		
Variable	n	Stable SCD, n = 291	Progressive SCD, n = 63	n	Stable SCD, n = 46	Progressive SCD, n = 5	n	Stable SCD, n = 186	Progressive SCD, n = 83
Demographics									
Female, n (%)*	354	138 (47)	26 (41)	51	34 (74)	4 (80)	269	71 (38)	34 (41)
Age in years	354	**61.2 ± 9.6**	**69.0 ± 7.1**	51	64.9 ± 6.4	70.2 ± 8.3	269	**64.5 ± 7.8**	**68.0 ± 8.4**
Education in years	354	13.3 ± 4.3	14.0 ± 4.4	51	10.8 ± 4.2	11.6 ± 4.3	269	**12.5 ± 2.8**	**13.3 ± 3.3**
Follow-up in years	354	**3.4 ± 2.2**	**3.8 ± 3.2**	51	3.7 ± 1.8	2.8 ± 1.8	269	**2.3 ± 0.9**	**1.6 ± 0.7**
MCI/AD/non-AD, n			42/15/6			2/2/1			53/21/9
APOE status									
APOE ε4 carrier, n (%)*	317	**92 (35)**	**27 (54)**	49	10 (22)	2 (50)	226	56 (35)	32 (47)
Neuropsychology									
MMSE	351	28.4 ± 1.7	28.0 ± 1.5	51	**28.3 ± 1.5**	**26.8 ± 1.9**	265	**28.2 ± 1.6**	**27.6 ± 1.8**
Memory, immediate recall	304	**41 ± 9**	**37 ± 8**	51	42 ± 5	38 ± 6	269	**20 ± 3**	**18 ± 4**
Memory, delayed recall	303	**8 ± 3**	**6 ± 3**	51	**14 ± 6**	**13 ± 2**	269	**7 ± 2**	**5 ± 2**
TMT-A, seconds	318	40 ± 19	44 ± 14	50	44 ± 16	47 ± 18	264	**42 ± 15**	**51 ± 20**
TMT-B, seconds	318	**97 ± 51**	**113 ± 48**	50	135 ± 87	163 ± 103	264	**102 ± 41**	**127 ± 52**
Category fluency	312	**22 ± 6**	**21 ± 5**	51	21 ± 5	17 ± 4	269	**21 ± 5**	**20 ± 5**
MRI									
Hippocampal volume, mL	332	**7.96 ± 0.83**	**7.49 ± 0.81**	49	8.20 ± 0.80	7.77 ± 1.12	93	**7.92 ± 0.84**	**7.19 ± 1.12**
cMTA	332	**0.37 ± 0.46**	**0.54 ± 0.54**	49	0.22 ± 0.43	0.40 ± 0.54	93	**0.54 ± 0.53**	**1.08 ± 0.86**
cGCA	332	0.75 ± 0.65	0.87 ± 0.62	49	0.10 ± 0.24	0.22 ± 0.36	93	**0.49 ± 0.64**	**1.17 ± 0.90**
Grading	332	**0.22 ± 0.19**	**0.36 ± 0.22**	49	0.09 ± 0.12	0.23 ± 0.22	93	**0.21 ± 0.23**	**0.44 ± 0.32**
CSF									
$A\beta_{42}$, pg/mL	227	**875 ± 235**	**638 ± 279**	41	771 ± 221	637 ± 194	87	**846 ± 300**	**670 ± 305**
Total tau, pg/mL	227	**266 ± 146**	**456 ± 370**	41	**333 ± 227**	**645 ± 694**	87	**286 ± 152**	**454 ± 281**
p-tau, pg/mL	227	**46 ± 18**	**65 ± 34**	41	55 ± 28	83 ± 65	87	**48 ± 20**	**63 ± 35**

Abbreviations: SCD, subjective cognitive decline; ADC, Amsterdam Dementia Cohort; DCN, Dementia Competence Network; AD, dementia due to Alzheimer's disease; FTD, frontotemporal dementia; VaD, vascular dementia; DLB, Lewy body dementia; MMSE, Mini–Mental State Examination; RAVLT, Rey Auditory Verbal Learning Task; FSCRT, Free and Cued Selective Reminding Test; CERAD, Consortium to Establish a Registry for Alzheimer's Disease; TMT, Trail Making Test; cGCA, computed cortical atrophy score, estimated using gray matter concentration; cMTA, computed medial temporal lobe atrophy score, (left + right)/2, derived from volumes of hippocampus and lateral ventricles; $A\beta_{42}$, amyloid-β 1-42; p-tau, tau phosphorylated at threonine 181.

NOTE. Follow-up in years: time to conversion to MCI/dementia or follow-up time for nonconverters. Non-AD cases consisted of (1) ADC: 3 FTD and 3 VaD; (2) Barcelona: 1 DLB; and (3) DCN: 1 FTD, 1 VaD, 3 DLB, and 4 nonspecified dementia.

NOTE. Memory, immediate recall: data on immediate recall using RAVLT (ADC), FSCRT (Barcelona), and CERAD (DCN); memory, delayed recall: data on delayed recall using RAVLT (ADC), FSCRT (Barcelona), and CERAD (DCN); hippocampal volume: left plus right hippocampus (in mL), normalized for head size and gender; grading: computed using a region of interest around the hippocampus, describing the intensity similarity of test image and training set images.

NOTE. Raw data are presented as mean ± SD or n (%). Group differences per center according to outcomes were calculated using Student's t-test for continuous variables. Bold represents P values $< .05$.

*For categorical variables, the chi-square test was used.

a computed medial temporal lobe atrophy (cMTA) score, a computed global cortical atrophy (cGCA) score, and region-of-interest (ROI)–based grading. They were derived as follows: first, whole-brain segmentation into 136 structures was performed using multi-atlas segmentation method [39]. From these structures, total (left + right) hippocampal volume was used in the classification. In addition, cMTA score was derived from the volumes of the hippocampus and inferior lateral ventricles [40]. Similarly, cGCA score was estimated using voxel-based morphometry [40]. Finally, the ROI-based grading method measures the similarity of the patient image to patient images from a certain diagnostic group. In practice, an ROI from the patient is represented as a linear combination of the corresponding ROIs from a database of reference images. As each reference image contains also information about the patient's diagnostic label, the grading feature is defined as the share of the weights from images with a certain diagnostic label. In this work, we used an ROI centered around the hippocampus [41]. See Supplementary Fig. A2, for a schematic presentation of this method. For classification, the volume of the hippocampus was normalized first for head size [42] and then for gender using the LMS method (referring to smooth curve [L], mean [M], and coefficient of variation [S]) [43]. The grading feature was also normalized for gender.

2.5. CSF

CSF samples from both ADC (n = 227) and Barcelona (n = 41) were analyzed at the Neurochemistry Laboratory at the Department of Clinical Chemistry of the VUmc, the Netherlands. In DCN (n = 87), samples were analyzed at the laboratory of the University of Erlangen, Germany. All centers measured amyloid-β 1-42 ($A\beta_{42}$), total tau, and tau phosphorylated at threonine 181 (p-tau) with commercially available ELISAs (Innotest; Fujirebio, Ghent, Belgium).

2.6. APOE genotyping

In ADC (n = 317), the apolipoprotein E (APOE) genotype was determined with the LightCycler APOE mutation detection method (Roche diagnostics GmbH, Mannheim, Germany). In Barcelona (n = 49), the APOE genotype was determined with PCR amplification and Sanger sequencing (ThermoFisher, USA). In DCN (n = 226), leukocyte DNA was isolated with the Qiagen Isolation Kit (Qiagen, Hilden, Germany). Patients were dichotomized into APOE ε4 carriers (heterozygous and homozygous) and noncarriers.

2.7. Disease State Index

For classifying patients at risk of progression or not, we used a modification of the PredictND tool that was previously developed in the European FP7 project PredictND (www.predictnd.eu). The tool is based on the DSI classifier [17]. When presented with a new patient, the DSI estimates the similarity of measurement values from this patient to observed values from reference patients with and without a certain medical condition, in this case similarity to patients with stable SCD and patients progressing to MCI or dementia [17]. Similarity is estimated in the following way: (1) Each measurement value of an individual person is compared with the reference data using a fitness function defined as $f(x) = FN(x)/(FN(x)+FP(x))$, where FN is the false-negative error rate, and FP is the false-positive error rate in the reference data when using the individual's measurement value x as a cutoff value in classification. (2) The "relevance" of each determinant is defined as sensitivity + specificity − 1. (3) Finally, a composite DSI is defined as a weighted average of fitness values: DSI = Σ (relevance • fitness)/Σ relevance. DSI is a continuous value between zero and one, reflecting how similar an individual is to patients who have previously progressed. A cutoff value of 0.5 is used to classify whether an individual patient is more likely to remain stable (DSI < 0.5) or progress to MCI or dementia (DSI ≥ 0.5) at follow-up. In addition, we studied whether the performance is improved for a subset of patients with high (DSI > 0.7 and DSI > 0.8) or low (DSI < 0.3 and DSI < 0.2) DSI values. This could enable detecting patients with very low risk or very high risk of progression for clinical counseling. The classifier also provides a visual representation of how different features contribute to the DSI in a so-called disease state fingerprint (see Fig. 2, for details). As the DSI combines multiple independent classifiers (fitness functions), there is no need to impute data or exclude cases with incomplete data. More mathematical details can be found in the study by Mattila et al. [17].

2.7.1. Development and internal validation

We developed the model on the ADC data and internally validated this model on the same cohort using 10 iterations of three-fold cross-validation. We assessed the different data sources separately (demographics, APOE status, neuropsychology tests, CSF biomarkers, and computed MRI imaging markers) and then combined them, independent of missing data. Owing to the technical differences between scanners, we excluded MRI features from patients scanned with 1.0 T devices (n = 183) from the training set and tested using all field strength and only >1.0 T. In this way, the classifier is able to better learn the differences between diagnostic groups without the excess variation from the scanner differences (for details, see Supplementary Table A1). We used the following performance metrics in the evaluation of the DSI: the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV). Although DSI balances the results by default, we also estimated balanced accuracy that is typically defined as mean of sensitivity and specificity. As an outcome measure, we defined progression to MCI or dementia, and we also repeated the analyses including only progression to MCI or dementia due to AD (excluding other dementias).

2.7.2. External validation

For external validation, we tested our developed model on new, unseen cases from pooled Barcelona and DCN data. To understand why the performance decreased with the independent validation cohort, we repeated the analyses by training the model and performing cross-validation with the Barcelona and DCN data, and using the ADC data as a separate validation cohort.

2.7.3. Comparison to other machine-learning algorithms

Earlier studies have performed comprehensive comparisons between the DSI classifier and other machine-learning algorithms [17,44,45]. We add on to this by comparing the classifier to Naïve Bayes and Random Forest classifiers. Details can be found in Appendix.

2.8. Other statistical analyses

We investigated differences in baseline characteristics in each center according to outcome, using Student's t-test and the chi-square test when appropriate, using SPSS, version 22 (IBM, Armonk, NY, USA). $P < .05$ was considered significant. The DSI analysis was performed using MATLAB toolbox in MATLAB, version R2015b (MathWorks, Natick, MA, USA) [46].

3. Results

3.1. Baseline characteristics

After a mean of 2.9 ± 2.0 years, 151 (22%) patients showed clinical progression to MCI or any type of dementia (Table 1). Patients who showed progression were older, more frequent APOE ε4 carriers, performed somewhat worse on

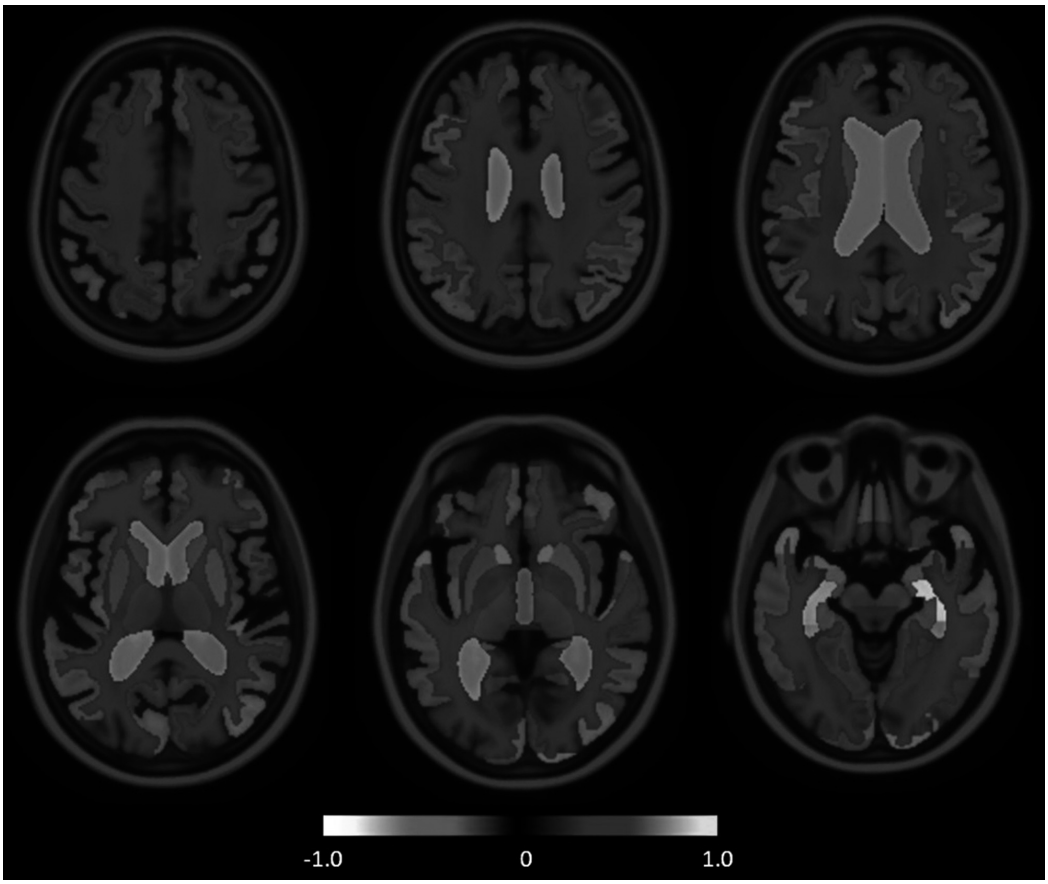


Fig. 1. The visualization of group-wise volume differences between stable subjective cognitive decline (SCD) and progressive SCD groups. The map visualizes the relative volume difference: $\frac{V_p - V_s}{0.5 \times (V_p + V_s)}$, where V_p and V_s are the mean volumes for progressive and stable groups, respectively. Blue indicates the structures on MRI that were larger in the progressive group, and red indicates the structures that were smaller.

neuropsychological tests, and had smaller hippocampal volumes and more abnormal CSF biomarkers. Patients in ADC were younger as compared with Barcelona and DCN. Patients in Barcelona were more often female, had less education, and showed less progression as compared with ADC and DCN. Duration of follow-up was longest in Barcelona and shortest in DCN. There were no differences in percentage of APOE ε4 carriers and baseline Mini–Mental State Examination across the centers.

3.2. Development and internal validation of the model

Table 2 shows performance of the DSI using the different data sources, for progression to MCI or dementia. As single data source, CSF showed highest balanced accuracy, followed by the automatic MRI features. Fig. 1 shows the group-wise volume differences between stable SCD and progressive SCD groups, with the clearest differences observed in the medial temporal region.

When we used all the data sources together, performance improved (balanced accuracy: 74.0 ± 5.5%). The model had high NPV (93.3 ± 2.8), whereas PPV was only modest (37.7 ± 5.5). This indicates that the DSI classifier was most useful to identify patients who remained stable. When we repeated the analyses for progression to MCI or dementia due to AD (excluding other dementias) as an outcome measure, results were comparable (Supplementary Table A2).

Table 2 also presents performance of the DSI classifier for subgroups having high or low DSI values, to aid the clinician on how to interpret the DSI values. We observed extreme DSI values, that is, below 0.3 or above 0.7, in 48 ± 6% of the patients. When DSI < 0.3, NPV was 97.0 ± 2.6, indicating that the probability of progression is very low in

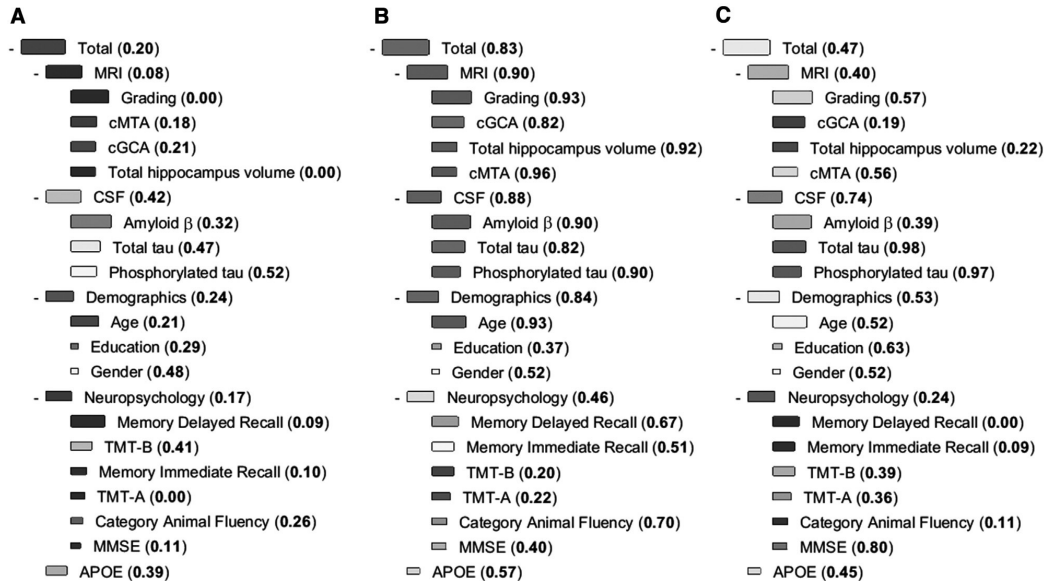


Fig. 2. Examples of DSI fingerprints: patient A and patient C remained stable, and patient B progressed to MCI. The DSI fingerprint combines all data available from one patient and displays it in a visually attractive format to the clinician. The DSI value is presented both numerically and visually with color. The color changes from blue to red when DSI increases from zero (high similarity to the stable group) to one (high similarity to the progressive group). The relevance is visualized by the size of the box. The larger the box, the better the specific marker discriminates the stable and progressive SCD patients. Abbreviations: MMSE, Mini–Mental State Examination; TMT, Trail Making Test; cGCA: computed cortical atrophy score, estimated using gray matter concentration; cMTA, computed medial temporal lobe atrophy score, (left + right)/2, derived from volumes of hippocampus and lateral ventricles; Amyloid β, amyloid-β 1–42; Phosphorylated tau, tau phosphorylated at threonine 181; DSI, Disease State Index.

this subset and the clinician could reassure these patients with high confidence. For comparison, if NPV is computed for all patients without using any prediction model, it is 82.0 [291/(291 + 63)], showing that DSI can clearly help in stratifying patients. When DSI $>$ 0.7, PPV was not very high, only 50.8 $\pm$ 12.9. Although the progression of an individual cannot be predicted accurately even in this subgroup, the risk of conversion is clearly elevated. The risk ratio is 2.8 in this subgroup compared with the whole patient population meaning that the clinician might start applying more rigorous follow-up and lifestyle intervention measures to these patients. This means that for roughly half of SCD patients, the DSI could have practical use to aid in individualized prognosis.

Fig. 2 shows the DSI fingerprints for three example patients to illustrate how the tool integrates and visualizes available data. Patient A is a 60-year-old female, with a DSI of 0.20, meaning the clinician can reassure her with high accuracy. Nearly all the boxes in the fingerprint are blue, which fits with the good outcome in this patient; she remained stable during three years of follow-up. Patient B is a 74-year-old female with a DSI value of 0.83, mainly attributable to her values on MRI and CSF (visible as red boxes). This implies her risk of progression is clearly elevated, and follow-up should be discussed. This patient progressed to MCI after 3 years. Patient C is a 66-year-old female, who remained stable during a follow-up period of 4 years. The fingerprint shows both red and blue boxes, implying that interpretation is inconclusive and a reliable prognosis cannot be made, further illustrated by an inconclusive DSI value of 0.47.

3.3. External validation

When we externally validated our model by testing it in pooled data of Barcelona and DCN, we found an overall lower performance (balanced accuracy 65.1, NPV 83.7; Table 3). Balanced accuracy increased to 78.5 in the more extreme DSI values. To evaluate what caused the lower performance on external validation, we also trained the model on pooled data of Barcelona and DCN and tested it in ADC data (Supplementary Table A3). Even when we developed the model in Barcelona and DCN cohorts, performance was still better in ADC (balanced accuracy 73.3, NPV 92.4).

3.4. Comparison to other machine-learning algorithms

For comparison, other machine-learning algorithms were also tested. The performance of the Naïve Bayes classifier was corresponding to and the Random Forest classifier lower than what was obtained by the DSI classifier (Supplementary Table A4).

Table 2
Performance of DSI to predict conversion to MCI or any type of dementia in the ADC cohort, for the total cohort and for patients with extreme DSI values

Variable	%	Stable SCD, n	Progressive SCD, n	AUC	Balanced accuracy	Sensitivity	Specificity	PPV	NPV
Demographics		291	63	0.74 ± 0.04	66.0 ± 5.0	66.0 ± 11.7	65.9 ± 6.1	29.7 ± 3.8	90.1 ± 2.8
APOE		267	50	0.60 ± 0.05	59.7 ± 4.9	53.9 ± 8.4	65.5 ± 4.5	22.7 ± 4.4	88.4 ± 2.4
Neuropsychology		290	62	0.69 ± 0.06	62.7 ± 4.4	61.6 ± 10.8	64.3 ± 5.3	26.9 ± 3.3	88.7 ± 2.4
CSF		194	33	0.77 ± 0.07	69.9 ± 5.0	66.1 ± 11.3	73.6 ± 6.5	30.3 ± 5.8	92.8 ± 2.6
MRI (1 T, 1.5 T, 3 T)		277	55	0.68 ± 0.05	61.4 ± 4.3	80.1 ± 10.1	42.8 ± 6.8	21.8 ± 2.5	91.9 ± 3.5
MRI (>1 T)		123	25	0.73 ± 0.09	69.1 ± 7.8	64.9 ± 15.2	73.3 ± 7.6	33.6 ± 9.6	91.3 ± 4.0
Demographics + APOE + Neuropsychology + CSF + MRI (1 T, 1.5 T, 3 T)		291	63	0.80 ± 0.05	74.0 ± 4.2	82.9 ± 8.4	65.1 ± 5.8	34.2 ± 3.8	94.7 ± 2.4
Demographics + APOE + Neuropsychology + CSF + MRI (>1 T)		291	63	0.81 ± 0.06	74.1 ± 5.8	75.7 ± 11.2	72.6 ± 4.8	37.7 ± 5.5	93.3 ± 2.8
DSI < 0.2 or DSI > 0.8									
Demographics + APOE + Neuropsychology + CSF + MRI (1 T, 1.5 T, 3 T)	14 ± 4	12 ± 5	5 ± 2	0.81 ± 0.10	83.3 ± 7.4	98.9 ± 4.2	67.7 ± 13.6	59.0 ± 17.4	99.4 ± 2.3
Demographics + APOE + Neuropsychology + CSF + MRI (>1 T)	21 ± 6	20 ± 8	5 ± 2	0.83 ± 0.11	84.1 ± 9.6	85.4 ± 17.6	82.8 ± 7.1	56.2 ± 17.1	96.3 ± 4.6
DSI < 0.3 or DSI > 0.7									
Demographics + APOE + Neuropsychology + CSF + MRI (1 T, 1.5 T, 3 T)	37 ± 6	34 ± 7	10 ± 3	0.84 ± 0.06	80.7 ± 6.0	89.6 ± 11.4	71.8 ± 9.4	47.8 ± 10.9	96.8 ± 3.0
Demographics + APOE + Neuropsychology + CSF + MRI (>1 T)	48 ± 6	47 ± 7	9 ± 3	0.84 ± 0.09	84.1 ± 7.3	84.9 ± 14.2	83.3 ± 5.5	50.8 ± 12.9	97.0 ± 2.6

Abbreviations: AUC, area under the receiver operating characteristic curve; SCD, subjective cognitive decline; PPV, positive predictive value; NPV, negative predictive value; APOE, apolipoprotein E; DSI, Disease State Index.

NOTE. For the extreme DSI values, n: number of patients in a cross-validation fold having the DSI value in the given range; %: percentage of patients in a test set (n = 118) of a cross-validation fold having the DSI value in the given range. Values are presented as mean ± standard deviation over 10 iterations of three-fold cross-validation.

Table 3
External validation: Performance of DSI to predict conversion to MCI or any type of dementia when tested in the pooled data of Barcelona and DCN cohorts, for the total cohort and for patients with extreme values

Variable	%	Stable SCD, n	Progressive SCD, n	AUC	Balanced accuracy	Sensitivity	Specificity	PPV	NPV
Demographics		232	88	0.63	57.8	61.4	54.3	33.8	78.8
APOE		203	72	0.57	57.4	47.2	67.5	34.0	78.3
Neuropsychology		232	88	0.69	63.9	63.6	64.2	40.3	82.3
CSF		90	39	0.69	61.7	59.0	64.4	41.8	78.4
MRI		100	42	0.77	67.4	73.8	61.0	44.3	84.7
Demographics + APOE + Neuropsychology + CSF + MRI		232	88	0.72	65.1	68.2	62.1	40.5	83.7
DSI < 0.2 or DSI > 0.8									
Demographics + APOE + Neuropsychology + CSF + MRI	21	38	30	0.81	78.5	83.3	73.7	71.4	84.8
DSI < 0.3 or DSI > 0.7									
Demographics + APOE + Neuropsychology + CSF + MRI	45	94	50	0.79	74.2	76.0	72.3	59.4	85.0

Abbreviations: AUC: area under the receiver operating characteristic curve; PPV, positive predictive value; NPV, negative predictive value; APOE, apolipoprotein E; DSI, Disease State Index; SCD, subjective cognitive decline.

NOTE. For the extreme DSI values, n: number of patients having the DSI value in the given range; %: percentage of patients having the DSI value in the given range. Values are presented as mean.

4. Discussion

In this large memory clinic study, we found that after an average follow-up of almost 3 years, 22% of the individuals with SCD showed clinical progression to MCI or dementia. The DSI classifier combining cognitive test results, automated MRI features, and CSF biomarkers accurately classified 74% of the patients, with especially high NPV. Nearly half of the patients had a clearly positive or negative DSI of <0.3 or >0.7, where balanced accuracy was as high as 84%.

Although many individuals with SCD may indeed be "worried well," a minority visits the memory clinic because they actually experience cognitive decline, which the clinician is not yet able to verify. We show that a computer-aided decision tool could support clinicians in identifying that minority of individuals who are at high risk of clinical progression. Moreover, for a larger group of individuals, reassurance can be even more explicit, backed up by negative findings on a combination of diagnostic tests. For daily clinical routine, this could imply a paradigm shift; it is current practice to reassure patients with SCD but not disclose results of their particular diagnostic tests. Our results provide support for the notion, however, that we approach an era of personalized medicine, where individuals' results on diagnostic tests can be used to obtain individualized predictions. Our classifier may aid in providing prognosis or decide to follow up individuals at increased risk for progression. On further scrutinizing the data, we observed that performance was particularly good for roughly half of the population, with a high or low DSI (<0.3 or >0.7), while prognostic performance was suboptimal for those with a medium DSI (0.4-0.6) (data not shown). Yet, overall NPV was very high, reaching up to 97.0 for the cases with DSI < 0.3. This implies that patients with a DSI < 0.3 can be reassured and do not need follow-up. For patients with a DSI > 0.7, a certain prognosis cannot be made, but the risk of clinical progression is clearly elevated and follow-up is warranted. The fingerprint could further aid in this interpretation by visualizing how each of the determinants contributed to the prognosis. Of note, in the present study, we focused on patients who present to a memory clinic with the clinical question whether they have an underlying neurodegenerative disease. In further work, tools like this could also be used for screening patients at risk in the general population, for example, by using blood-based biomarkers [47].

The overall balanced accuracy of the DSI was highest when we combined all different data sources. The discriminative effect of MRI and CSF biomarkers are in line with the additive model, indicating patients with SCD at risk of progression already have more AD-like biomarkers at baseline [48]. Also, neuropsychological assessment at baseline improved the performance of the DSI. It is conceivable that even within normal boundaries, a slight decline in cognitive performance is associated with progression, which is particularly appreciated when analyzed together with data from other sources. The classifier also provided fully automatically computed MRI features enabling the clinician to extract more information from the images than when using visual interpretation only [19].

The strength of this study was the large size of the cohort in which the model was developed, and the availability of two independent cohorts for external validation. All patients underwent thorough examination and were only included if cognitive complaints could not be confirmed by cognitive testing. We used data that were typical of memory clinics, varied and incomplete. Because we aimed to develop a tool that should be able to support clinicians in daily practice, it is essential the tool can deal with missing data.

However, several potential limitations also need to be discussed. In general, when developing prediction models based on classifiers, comparing training and testing results can be challenging for several reasons. In this study we trained the tool on the ADC data and found that on validation in the Barcelona and DCN data, performance was less optimal. This might indicate that generalizability is limited. When we trained the tool in the Barcelona and DCN data and then performed external validation in ADC, we still found that performance was better in ADC than in the Barcelona and DCN data. This suggests that not the model itself hampers generalizability, but the lower performance is caused by heterogeneity in cohorts. Overall, the following sources can affect generalizability of prediction models: (1) patients in different memory clinics are different (i.e., both referral and definition of SCD), (2) heterogeneity in outcome, (3) patient measurements are done in different ways, and (4) prediction models are not able to generalize. In the field of SCD, heterogeneity between cohorts is an important hurdle [2,5,49]. The field is acknowledging this and working toward more harmonization of research efforts. Nonetheless, it is of the utmost importance to actually perform studies on multiple data sets, both to get to know the differences and how this influences results, and to start harmonizing and bridging data. In this study, we feel there are several important cohort differences: first, patients showed substantial baseline differences between the three memory clinic cohorts. We found differences regarding progression rates and definition of progression; ADC and Barcelona used the Petersen criteria for MCI, whereas DCN used the Jak-Bondi criteria for MCI [28,30]. Also those who remained stable in Barcelona and DCN were older than those in ADC. Second, although follow-up duration in VUmc was longer, more patients showed progression in Barcelona + DCN. Third, although all patients underwent a harmonized work-up, the work-up differed between the centers. We tried to eliminate these differences as much as possible. For the neuropsychology tests, we selected tests that overlapped or resembled each

other. Also, CSF analyses of ADC, DCN, and Barcelona were performed in the two laboratories, as part of the Euro-SCD collaboration, minimizing, but not excluding, interlaboratory variability. MRI scans were acquired on systems with different field strengths, yet the automatic analyses of these scans were all performed by the same software [19]. However, 1.0 T images have worse gray matter–white matter contrast than 1.5 T and 3.0 T images. Consequently, we decided to use only 1.5 T and 3.0 T images in training to have a robust classifier and then reported the results separately for different field strengths to demonstrate the differences between 1.0 T and >1.0 T images with roughly similar performance. In this study, we did not perform feature selection and choose a set of features maximizing prediction accuracy. We included diagnostics tests and features that are either familiar to clinicians or which we found to be good features in other studies. Had we used an optimal set of features, this would probably increase the performance of our model, at the risk of overfitting.

In conclusion, this study shows that it is feasible to extract and combine information from routine diagnostic tests into a measure that can be used within a clinical decision support system, supporting clinicians to identify individuals at risk of progression who need follow-up and individuals who are likely to remain stable and can be reassured and discharged. This implies that it is possible to think about a personalized medicine approach, also in patients with SCD. Recent research has shown that patients would like to be actively involved in decisions about prognostic testing, but they feel they often lack important information on the implication of the tests [15,50]. Tools such as the DSI classifier can provide a first step in taking personalized medicine in SCD to a next level.

Acknowledgments

Research of the VUmc Alzheimer Center is part of the neurodegeneration research program of the Amsterdam Neuroscience. The VUmc Alzheimer Center is supported by Alzheimer Nederland and Stichting VUmc Fonds. The clinical database structure was developed with funding from Stichting Dioraphte. For the development of the PredictAD tool, the VTT Technical Research Centre of Finland has received funding from European Union's Seventh Framework Programme for research, technological development, and demonstration under grant agreements 601055 (VPH-DARE@IT), 224328 (PredictAD), and 611005 (PredictND). The Euro-SCD project has been funded by the EU Joint Program–Neurodegenerative Disease Research (JPND_PS_FP-689-019). DCN has been funded by a grant from the German Federal Ministry of Education and Research (BMBF): Kompetenznetz Demenzen (01GI0420). Hanneke FM Rhodius-Meester is appointed on PredictND, a grant from the European Seventh Framework Program project PredictND under grant agreement 611005. Frederik Barkhof is supported by the NIHR UCLH Biomedical Research Center. Sietske AM Sikkes is supported by an Off Road grant (ZonMw #451001010). Wiesje M. van der Flier is a recipient of a research grant from Gieskes-Strijbis Fonds. Betty M. Tijms receives grant support from ZonMw (#73305056 and #733050824).

Author disclosures: Hanneke F.M. Rhodius-Meester, Hilkka Liedes, Steffen Wolfsgruber, Nina Coll-Padros, Johannes Kornhuber, Luca Kleineidam, Lorena Rami, Sietske A. Sikkes, Linda MP Wesselman, Rosalinde E.R. Slot, Sander C.J. Verfaillie, and Betty Tijms report no disclosures. Juha Koikkalainen and Jyrki Lötjönen all report that the VTT Technical Research Center of Finland owns the following IPR related to the article: (1) J. Koikkalainen and J. Lotjonen—a method for inferring the state of a system, US7,840,510 B2, PCT/FI2007/050277; and (2) J. Lotjonen, J. Koikkalainen, and J. Mattila—state inference in a heterogeneous system, PCT/FI2010/050545, FI20125177. Juha Koikkalainen and Jyrki Lötjönen are shareholders in Combinostics Oy. Oliver Peters has received speaker honoraria from Eli Lilly & Company, Novartis, Affiris, and Roche and has received research support from Axon, Axovant, Biogen, Eli Lilly and Company, Lundbeck, Pharmatrophix, Probiodrug, Novartis, Roche, Janssen, Piramal, Takeda, and TRX Pharmaceuticals. Frank Jessen has received fees for advisory boards of Eli Lilly, Biogene, MSD, Janssen Cilag, Roche, AC Immune, and Novartis. José Luis Molinuevo is the PI of trials funded by Roche, Merck, Novartis, and Janssen and has received speaker or consultant fees from Roche, Roche diagnostics, Biogen, Merck, Novartis, Oryzon, IBL, Axovant, Lundbeck, and Lilly. Charlotte E. Teunissen is a member of the Innogenetics International Advisory Boards of Fujirebio/Innogenetics and Roche. Frederik Barkhof serves/has served on the advisory boards of Bayer-Schering Pharma, Sanofi-Aventis, Biogen-Idec, TEVA, Merck-Serono, Novartis, Roche, Synthon BV, Jansen Research, and Genzyme. He received funding from the Dutch MS Society and EU-FP7 and has been a speaker at symposia organized by the Serono Symposia Foundation and Medscape. Philip Scheltens has served as a consultant for Wyeth-Elan, Genentech, Danone, and Novartis and received funding for travel from Pfizer, Elan, Janssen, and Danone Research. Wiesje M. van der Flier performs contract research for Biogen. Research programs of Wiesje M. van der Flier have been funded by ZonMw, NWO, EU-FP7, Alzheimer Nederland, CardioVascular Onderzoek Nederland, Stichting Dioraphte, Gieskes-Strijbis Fonds, Boehringer Ingelheim, Piramal Neuroimaging, Combinostics, Roche BV, and Janssen Stellar. She has been an invited speaker at Boehringer Ingelheim and Biogen. All funding is paid to her institution.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dadm.2018.09.001.

RESEARCH IN CONTEXT

1. Systematic review: An increasing number of studies focus on biomarkers that can help identifying patients with subjective cognitive decline (SCD) at risk of progression. Translation to clinical practice is hampered because it remains unclear what the diagnostic workup and follow-up for SCD should look like and what results should be disclosed in daily practice. We cited relevant citations.
2. Interpretation: We used a clinical decision support system to identify patients with SCD at risk for progression. Clinical decision support systems can weigh and combine different diagnostic tests; this multivariate model showed especially a high negative predictive value, meaning the classifier identified patients who will remain stable and can thus be reassured.
3. Future directions: Clinical decision support systems could be useful to aid clinicians in interpreting diagnostic test results and discuss results of these tests with patients with SCD. To take diagnosis and prognosis in SCD to the next level, further knowledge on shared decision-making in SCD is needed.

References

[1] Buckley RF, Villemagne VL, Masters CL, Ellis KA, Rowe CC, Johnson K, et al. A Conceptualization of the utility of subjective cognitive decline in clinical trials of preclinical Alzheimer's disease. J Mol Neurosci 2016;60:354–61.

[2] Jessen F, Amariglio RE, van Boxtel M, Breteler M, Ceccaldi M, Chetelat G, et al. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. Alzheimers Dement 2014;10:844–52.

[3] van Harten AC, Visser PJ, Pijnenburg YA, Teunissen CE, Blankenstein MA, Scheltens P, et al. Cerebrospinal fluid Abeta42 is the best predictor of clinical progression in patients with subjective complaints. Alzheimers Dement 2013;9:481–7.

[4] Cavedo E, Lista S, Khachaturian Z, Aisen P, Amouyel P, Herholz K, et al. The Road Ahead to Cure Alzheimer's Disease: Development of Biological Markers and Neuroimaging Methods for Prevention Trials Across all Stages and Target Populations. J Prev Alzheimers Dis 2014;1:181–202.

[5] Molinuevo JL, Rabin LA, Amariglio R, Buckley R, Dubois B, Ellis KA, et al. Implementation of subjective cognitive decline criteria in research studies. Alzheimers Dement 2017;13:296–311.

[6] Fonseca JA, Ducksbury R, Rodda J, Whitfield T, Nagaraj C, Suresh K, et al. Factors that predict cognitive decline in patients with subjective cognitive impairment. Int Psychogeriatr 2015;27:1671–7.

[7] Toledo JB, Bjerke M, Chen K, Rozycki M, Jack CR Jr, Weiner MW, et al. Memory, executive, and multidomain subtle cognitive impairment: clinical and biomarker findings. Neurology 2015;85:144–53.

[8] Hessen E, Nordlund A, Stalhammar J, Eckerstrom M, Bjerke M, Eckerstrom C, et al. T-Tau is associated with objective memory decline over two years in persons seeking help for subjective cognitive decline: A report from the Gothenburg-Oslo MCI Study. J Alzheimers Dis 2015;47:619–28.

[9] Toledo JB, Weiner MW, Wolk DA, Da X, Chen K, Arnold SE, et al. Neuronal injury biomarkers and prognosis in ADNI subjects with normal cognition. Acta Neuropathol Commun 2014;2:26.

[10] Meiberth D, Scheef L, Wolfsgruber S, Boecker H, Block W, Traber F, et al. Cortical thinning in individuals with subjective memory impairment. J Alzheimers Dis 2015;45:139–46.

[11] van der Flier WM, van Buchem MA, Weverling-Rijnsburger AW, Mutsaers ER, Bollen EL, Admiraal-Behloul F, et al. Memory complaints in patients with normal cognition are associated with smaller hippocampal volumes. J Neurol 2004;251:671–5.

[12] Verfaillie SC, Tijms B, Versteeg A, Benedictus MR, Bouwman FH, Scheltens P, et al. Thinner temporal and parietal cortex is related to incident clinical progression to dementia in patients with subjective cognitive decline. Alzheimers Dement (Amst) 2016;5:43–52.

[13] Peter J, Scheef L, Abdulkadir A, Boecker H, Heneka M, Wagner M, et al. Gray matter atrophy pattern in elderly with subjective memory impairment. Alzheimers Dement 2014;10:99–108.

[14] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. 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. Alzheimers Dement 2011;7:280–92.

[15] Van der Flier W, Kunneman M, Bouwman FH, Petersen R, Smets EMA. Diagnostic dilemmas in Alzheimer's disease: room for shared decision making. Alzheimers Dement 2017;3:301–4.

[16] Frisoni GB, Boccardi M, Barkhof F, Blennow K, Cappa S, Chiotis K, et al. Strategic roadmap for an early diagnosis of Alzheimer's disease based on biomarkers. Lancet Neurol 2017;16:661–76.

[17] Mattila J, Koikkalainen J, Virkki A, Simonsen A, van GM, Waldemar G, et al. A disease state fingerprint for evaluation of Alzheimer's disease. J Alzheimers Dis 2011;27:163–76.

[18] Mattila J, Soininen H, Koikkalainen J, Rueckert D, Wolz R, Waldemar G, et al. Optimizing the diagnosis of early Alzheimer's disease in mild cognitive impairment subjects. J Alzheimers Dis 2012; 32:969–79.

[19] Koikkalainen J, Rhodius-Meester H, Tolonen A, Barkhof F, Tijms B, Lemstra AW, et al. Differential diagnosis of neurodegenerative diseases using structural MRI data. Neuroimage Clin 2016;11:435–49.

[20] Rhodius-Meester HF, Koikkalainen J, Mattila J, Teunissen CE, Barkhof F, Lemstra AW, et al. Integrating biomarkers for underlying Alzheimer's Disease in mild cognitive impairment in daily practice: Comparison of a clinical decision support system with individual biomarkers. J Alzheimers Dis 2015;50:261–70.

[21] Munoz-Ruiz MA, Hartikainen P, Hall A, Mattila J, Koikkalainen J, Herukka SK, et al. Disease state fingerprint in frontotemporal degeneration with reference to Alzheimer's disease and mild cognitive impairment. J Alzheimers Dis 2013;35:727–39.

[22] van der Flier WM, Pijnenburg YA, Prins N, Lemstra AW, Bouwman FH, Teunissen CE, et al. Optimizing patient care and research: the Amsterdam Dementia Cohort. J Alzheimers Dis 2014; 41:313–27.

[23] Slot RER, Verfaillie SCJ, Overbeek JM, Timmers T, Wesselman LMP, Teunissen CE, et al. Subjective Cognitive Impairment Cohort (SCIENCe): study design and first results. Alzheimer Res Ther 2018. In press.

[24] van der Flier WM, Scheltens P. Amsterdam Dementia Cohort: Performing Research to Optimize Care. J Alzheimers Dis 2018; 62:1091–111.

[25] Valech N, Mollica MA, Olives J, Tort A, Fortea J, Lleo A, et al. Informants' perception of subjective cognitive decline helps to discriminate preclinical Alzheimer's disease from normal aging. J Alzheimers Dis 2015;48:S87–98.

[26] Wolfsgruber S, Polcher A, Koppara A, Kleineidam L, Frolich L, Peters O, et al. Cerebrospinal fluid biomarkers and clinical progression in patients with subjective cognitive decline and mild cognitive impairment. J Alzheimers Dis 2017;58:939–50.
[27] Kornhuber J, Schmidtke K, Frolich L, Perneczky R, Wolf S, Hampel H, et al. Early and differential diagnosis of dementia and mild cognitive impairment: design and cohort baseline characteristics of the German Dementia Competence Network. Dement Geriatr Cogn Disord 2009;27:404–17.
[28] Petersen RC. Mild cognitive impairment as a diagnostic entity. J Intern Med 2004;256:183–94.
[29] Albert MS, Dekosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011; 7:270–9.
[30] Bondi MW, Edmonds EC, Jak AJ, Clark LR, Delano-Wood L, McDonald CR, et al. Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. J Alzheimers Dis 2014;42:275–89.
[31] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011;7:263–9.
[32] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 1984;34:939–44.
[33] Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12:189–98.
[34] Van der Elst W, Van Boxtel MP, Van Breukelen GJ, Jolles J. Normative data for the animal, profession and letter M naming verbal fluency tests for Dutch speaking participants and the effects of age, education, and sex. J Int Neuropsychol Soc 2006;12:80–9.
[35] Reitan R. Validity of the Trail Making Test as an indicator of organic brain damage. Percept Mot Skills 1958;8:271–6.
[36] Saan RD, BG. De 15-Woorden Test A en B. Een voorlopige handleiding. (in Dutch). Groningen: Afdeling Neuropsychologie, AZG; 1986.
[37] Grober E, Buschke H, Crystal H, Bang S, Dresner R. Screening for dementia by memory testing. Neurology 1988;38:900–3.
[38] Berres M, Monsch AU, Bernasconi F, Thalmann B, Stahelin HB. Normal ranges of neuropsychological tests for the diagnosis of Alzheimer's disease. Stud Health Technol Inform 2000;77:195–9.
[39] Lotjonen JM, Wolz R, Koikkalainen JR, Thurfjell L, Waldemar G, Soininen H, et al. Fast and robust multi-atlas segmentation of brain magnetic resonance images. Neuroimage 2010;49:2352–65.
[40] Lotjonen J, Koikkalainen J, Rhodius- Meester HFM, van der Flier WM, Scheltens P, Barkhof F, et al. Computed rating scales for cognitive disorders from MRI. Alzheimers Dement 2017;13:1108.
[41] Tong T, Wolz R, Coupe P, Hajnal JV, Rueckert D. Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling. Neuroimage 2013;76:11–23.
[42] Buckner RL, Head D, Parker J, Fotenos AF, Marcus D, Morris JC, et al. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. Neuroimage 2004;23:724–38.
[43] Cole TJ, Green PJ. Smoothing reference centile curves: the LMS method and penalized likelihood. Stat Med 1992;11:1305–19.
[44] Mattila J, Koikkalainen J, Virkki A, van GM, Lotjonen J. Design and application of a generic clinical decision support system for multiscale data. IEEE Trans Biomed Eng 2012;59:234–40.
[45] Tolonen A, Rhodius-Meester HFM, Bruun M, Koikkalainen J, Barkhof F, Lemstra AW, et al. Data-driven differential diagnosis of dementia using multiclass disease state index classifier. Front Aging Neurosci 2018;10:111.
[46] Cluitmans L, Mattila J, Runtti H, van Gils M, Lotjonen J. A MATLAB toolbox for classification and visualization of heterogenous multiscale human data using the Disease State Fingerprint method. Stud Health Technol Inform 2013;189:77–82.
[47] Verberk I, Slot RE, Verfaillie SC, Heijst H, Prins ND, Van Berckel B, et al. Plasma-amyloid as pre-screener for the earliest Alzheimer's pathological changes. Ann Neurol 2018. in press.
[48] Jack CR Jr, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. Lancet Neurol 2013;12:207–16.
[49] Slot RE, Sikkes S, Berkhof J, Brodaty H, Buckley R, Cavedo E, et al. Subjective cognitive decline and rates of incident Alzheimer's disease (AD) and non-AD dementia. Alzheimers Dement 2018. in press.
[50] Kunneman M, Pel-Littel R, Bouwman FH, Gillissen F, Schoonenboom NSM, Claus JJ, et al. Patients' and caregivers' views on conversations and shared decision making in diagnostic testing for Alzheimer's disease: the ABIDE project. Alzheimer's Dementia (N Y) 2017;3:314–22.

APPENDIX – Supplemental data

Table A.1 Baseline differences in automatic MRI features according to outcome, for ADC subjects only, stratifying for field strength.

	n	Stable SCD	Progr SCD	*p-value*
Total hippocampus volume, all MRI	332	7.96± 0.83	7.49± 0.81	0.000
Total hippocampus volume, >1T MRI	148	8.19± 0.83	7.67± 0.87	0.005
Total hippocampus volume, 1T MRI	184	7.78± 0.78	7.34± 0.73	0.005
cMTA, all MRI	332	0.37± 0.46	0.54± 0.54	0.020
cMTA, >1T MRI	148	0.30± 0.44	0.50± 0.54	0.047
cMTA, 1T MRI	184	0.43± 0.47	0.56± 0.55	0.163
cGCA, all MRI	332	0.75± 0.65	0.87± 0.62	0.222
cGCA, >1T MRI	148	0.31± 0.55	0.67± 0.73	0.006
cGCA, 1T MRI	184	1.10± 0.50	1.04± 0.47	0.493
Grading, all MRI	332	0.22± 0.19	0.36± 0.22	0.000
Grading, >1T MRI	148	0.13± 0.16	0.31± 0.22	0.000
Grading, 1T MRI	184	0.29± 0.19	0.41± 0.21	0.002

>1T: using only 1.5T and 3T images; 1T: using only 1T images; hippocampal volume: in millilitres (ml) left plus right hippocampus, normalized for head size and gender; cMTA: computed medial temporal lobe atrophy score, (left+right)/2, derived from volumes of hippocampus and lateral ventricles; cGCA: computed cortical atrophy score, estimated using grey matter concentration; grading: computed using a region of interest around the hippocampus, describing the intensity similarity of test image and training set images.

Raw data are presented as mean ± SD. Group differences were calculated using Student's t test.

Table A.2 Performance of DSI to predict conversion to MCI or AD in the ADC cohort.

Variable	Stable SCD n	Progr SCD n	AUC	Balanced accuracy	Sensitivity	Specificity	PPV	NPV
Demographics	291	57	0.73 ± 0.04	66.2 ± 4.9	66.5 ± 11.1	65.9 ± 5.2	27.7 ± 3.4	91.1 ± 2.4
APOE	267	45	0.62 ± 0.06	61.5 ± 6.5	57.5 ± 12.6	65.5 ± 3.3	21.9 ± 4.5	90.2 ± 2.6
Neuropsychology	290	57	0.70 ± 0.07	63.9 ± 6.2	62.8 ± 12.8	65.0 ± 3.8	26.0 ± 4.1	90.0 ± 3.0
CSF	194	29	0.79 ± 0.07	72.8 ± 7.7	70.0 ± 12.8	75.6 ± 7.0	30.9 ± 7.5	94.3 ± 2.6
MRI (1T, 1.5T, 3T)	277	50	0.68 ± 0.05	62.0 ± 4.0	78.8 ± 9.0	45.3 ± 6.0	20.6 ± 2.0	92.4 ± 3.1
MRI (>1T)	123	23	0.75 ± 0.09	68.4 ± 7.1	62.8 ± 17.1	74.1 ± 8.9	31.9 ± 8.4	91.6 ± 3.6
Demographics + APOE + Neuropsychology + CSF + MRI (1T, 1.5T, 3T)	291	57	0.81 ± 0.04	75.2 ± 3.8	84.2 ± 6.6	66.2 ± 5.8	33.1 ± 3.8	95.6 ± 1.8
Demographics + APOE + Neuropsychology + CSF + MRI (>1T)	291	57	0.81 ± 0.04	75.8 ± 4.1	77.5 ± 8.1	74.0 ± 4.3	37.2 ± 4.2	94.4 ± 1.9

AUC: area under the receiver operating characteristic curve;, PPV: positive predictive value; NPV: negative predictive value**.** Values are presented as mean ± standard deviation over ten iterations of three-fold cross validation. Subjects that progressed to other dementias were excluded from this analysis.

Table A.3 Generalizability of DSI to predict conversion to MCI or any type of dementia.

Variable	Stable SCD n	Progr SCD n	AUC	Balanced accuracy	Sensitivity	Specificity	PPV	NPV
1) Trained and tested in pooled data of Barcelona and DCN cohorts, using cross validation.								
Demographics	232	88	0.64 ± 0.05	61.6 ± 5.1	60.7 ± 10.7	62.6 ± 7.3	38.3 ± 4.9	80.9 ± 3.8
APOE	203	72	0.57 ± 0.05	57.3 ± 5.1	47.1 ± 7.9	67.5 ± 4.6	34.0 ± 5.6	78.2 ± 3.8
Neuropsychology	232	88	0.68 ± 0.05	63.5 ± 4.4	63.5 ± 8.1	63.6 ± 4.2	39.8 ± 4.1	82.2 ± 3.3
CSF	90	39	0.69 ± 0.09	62.8 ± 7.6	56.1 ± 15.6	69.6 ± 9.2	44.5 ± 11.8	79.0 ± 6.1
MRI (>1T)	100	42	0.77 ± 0.07	71.3 ± 6.5	69.7 ± 12.6	72.9 ± 7.1	52.0 ± 10.8	85.3 ± 5.9
Demographics + APOE + Neuropsychology + CSF + MRI (>1T)	232	88	0.73 ± 0.05	68.1 ± 4.4	66.8 ± 7.7	69.4 ± 5.8	45.6 ± 5.0	84.7 ± 3.1
2) Tested in ADC cohort.								
Demographics	291	63	0.69	66.2	69.8	62.5	28.8	90.5
APOE	267	50	0.60	59.8	54.0	65.5	22.7	88.4
Neuropsychology	290	62	0.68	63.2	61.3	65.2	27.3	88.7
CSF	194	33	0.78	72.8	63.6	82.0	37.5	93.0
MRI (1T. 1.5T. 3T)	277	55	0.66	60.4	69.1	51.6	22.1	89.4
MRI (>1T)	123	25	0.73	71.8	64.0	79.7	39.0	91.6
Demographics + APOE + Neuropsychology + CSF + MRI (1T, 1.5T, 3T)	291	63	0.79	72.7	74.6	70.8	35.6	92.8
Demographics + APOE + Neuropsychology + CSF + MRI (>1T)	291	63	0.81	73.3	71.4	75.3	38.5	92.4

AUC: area under the receiver operating characteristic curve; PPV: positive predictive value; NPV: negative predictive value. 1) tested in pooled data of Barcelona and DCN cohorts (values are presented as mean ± standard deviation over ten iterations of three-fold cross validation) and 2) trained in pooled data of the Barcelona and DCN cohort and tested in ADC cohort.

Appendix Comparison to other machine learning algorithms

Earlier studies have performed comprehensive comparisons between the DSI classifier and other machine learning algorithms [1-3]. These studies showed that the DSI is not always the best classifier for different datasets, but gives robustly good results. Furthermore the DSI has several benefits: 1) it has a graphical counterpart which makes interpretation of results to clinicians more transparent, 2) it tolerates missing data (no imputation needed), 3) is not just a dichotomous classifier but gives also information about the confidence of the classification and 4) it is not very prone to overlearning as no high-dimensional decision boundaries are defined (classifier defined independently for each feature).

In this study we compared the DSI classifier to Naïve Bayes (NB) and Random Forest (RF)classifiers. We repeated the analyses described in section 2.7.1 and 2.7.2, now using the NB and RF classifiers. First, we applied cross-validation on the ADC data and, subsequently, trained the model on ADC data and tested it on the pooled Barcelona and DCN data, taking conversion to MCI or dementia as our outcome measure. Many classifiers, like NB and RF, do not perform well when the class distributions are imbalanced, i.e. the number of positive cases differs greatly from the number of negative cases. In our data, the number of stable SCD subjects was greater than the number of progressors, thus, we 1) used the original data sets as such, and 2) increased the number of progressors to match the number of stable cases by adding data from randomly selected progressors multiple times.

Overall, performance of the NB classifier was corresponding and the RF classifier considerable worse than the DSI classifier. Details can be found in table A.4.

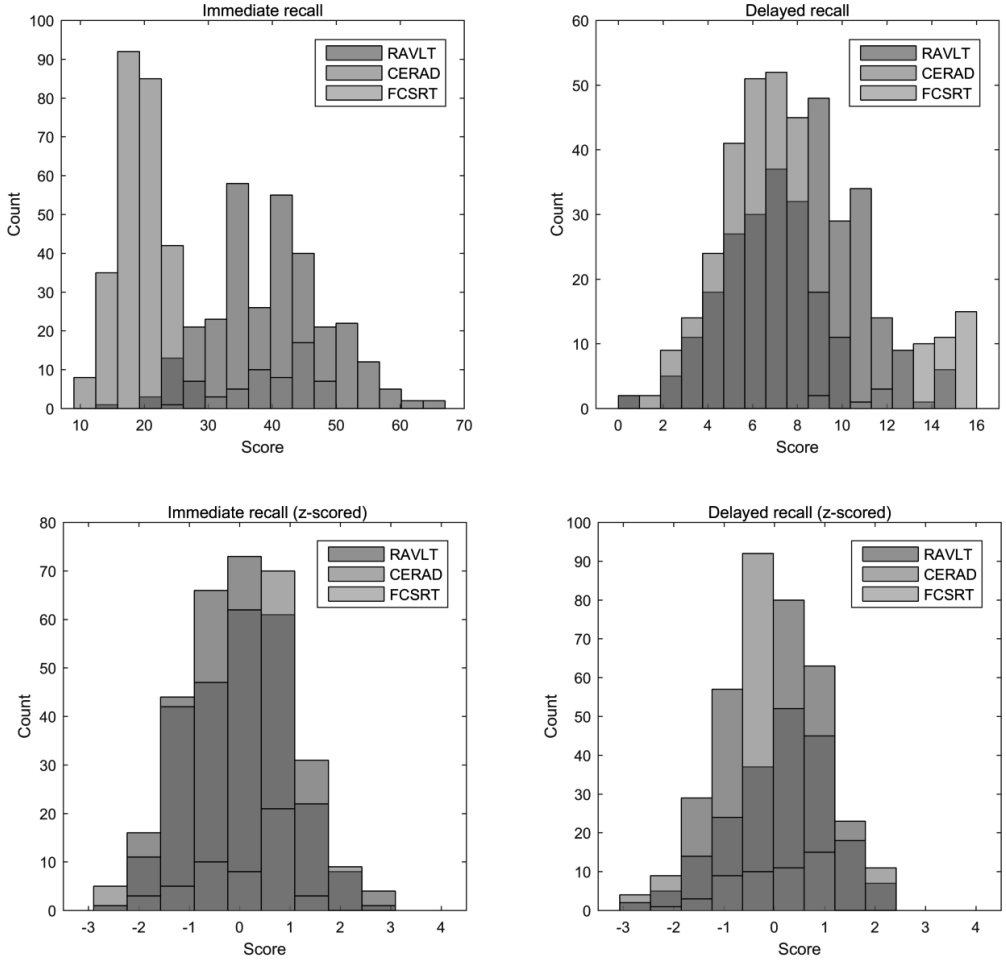
[1] Mattila J, Koikkalainen J, Virkki A, Simonsen A, van GM, Waldemar G, et al. A disease state fingerprint for evaluation of Alzheimer's disease. J Alzheimers Dis. 2011;27:163-76.
[2] Mattila J, Koikkalainen J, Virkki A, van GM, Lotjonen J. Design and application of a generic clinical decision support system for multiscale data. IEEE Trans Biomed Eng. 2012;59:234-40.
[3] Tolonen A, F. M. Rhodius-Meester H, Bruun M, Koikkalainen J, Barkhof F, Lemstra A, et al. Data-Driven Differential Diagnosis of Dementia Using Multiclass Disease State Index Classifier. Front Aging Neurosci. 2018;10:111.

Table A.4 Performance of Naïve Bayes (NB) and Random Forest (RF) classifier to predict conversion to MCI or any type of dementia, using all variables.

Variable	Stable SCD n	Progr SCD n	AUC	Balanced accuracy	Sensitivity	Specificity	PPV	NPV
1) Trained and tested in ADC cohort, using cross validation.								
RF classifier	291	63	0.76 ± 0.06	53.0 ± 2.4	7.3 ± 5.6	98.7 ± 1.4	60.6 ± 29.1	83.1 ± 0.8
NB classifier	291	63	0.78 ± 0.05	68.2 ± 5.2	54.6 ± 11.4	81.8 ± 5.8	40.4 ± 8.2	89.3 ± 2.2
RF classifier (increased number of progressors)	291	291	0.76 ± 0.06	67.6 ± 6.4	49.6 ± 11.2	85.6 ± 6.0	77.1 ± 9.5	63.4 ± 5.1
NB classifier (increased number of progressors)	291	291	0.80 ± 0.05	72.6 ± 5.1	71.0 ± 11.3	74.2 ± 6.4	73.3 ± 5.7	72.9 ± 7.1
2) Trained in ADC cohort and tested in pooled data of Barcelona and DCN cohorts.								
RF classifier	232	88	0.69	55.2	12.5	97.8	68.8	74.7
NB classifier			0.71	67.2	51.1	83.2	53.6	81.8
RF classifier (increased number of progressors)	232	232	0.66	62.1	47.4	76.7	67.1	59.3
NB classifier (increased number of progressors)	232	232	0.71	67.2	65.9	68.5	67.7	66.8

AUC: area under the receiver operating characteristic curve; PPV: positive predictive value; NPV: negative predictive value, RF: Random Forest, NB: Naïve Bayes. 1) trained and tested in ADC data (values are presented as mean ± standard deviation over ten iterations of three-fold cross validation) and 2) trained in ADC cohort and tested in pooled data of the Barcelona and DCN cohort, using DSI, NB and RF (number of trees in RF=100).

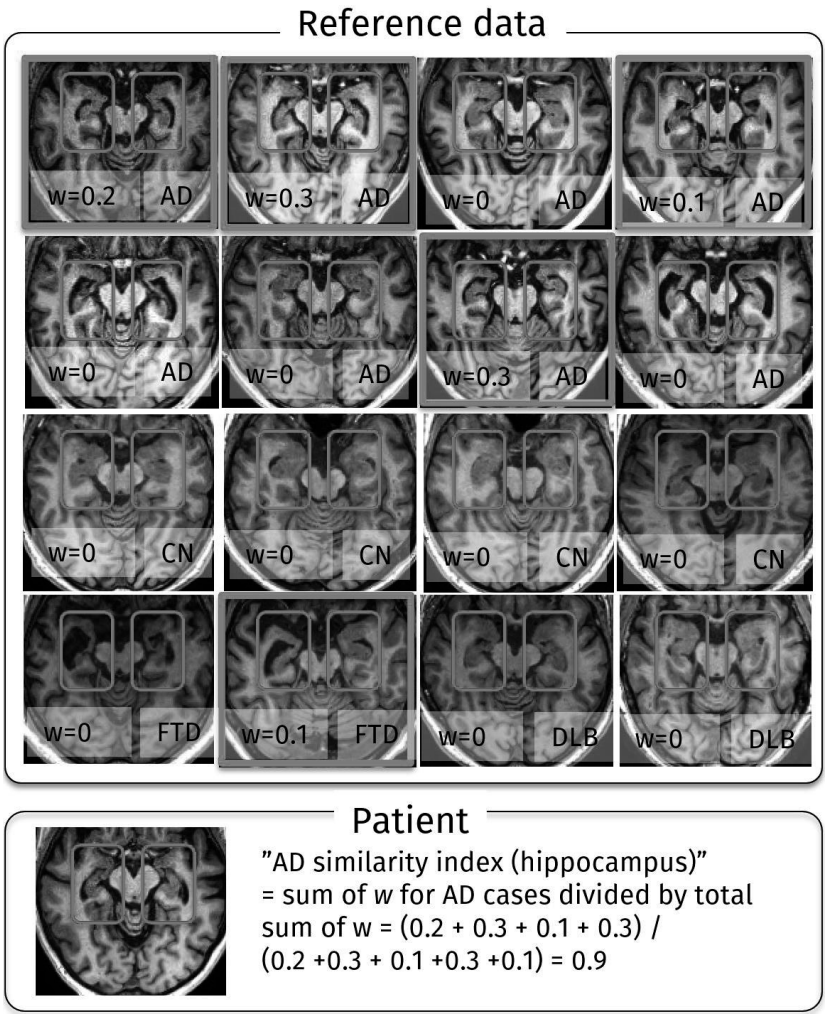
Figure A.1 Distribution of raw and normalized (z-scored) memory tests, with accompanying two- sample Kolmogorov-Smirnov test.



RAVLT: Rey auditory verbal learning task, CERAD: Consortium to Establish a Registry for Alzheimer's Disease, FSCRT: Free and Cued Selective Reminding test.

Two-sample Kolmogorov-Smirnov test		***p-value***	
		Raw scores	**Normalized**
RAVLT Immediate Recall	CERAD Immediate Recall	<0.00001	0.152
RAVLT Immediate Recall	FCSRT Immediate Recall	<0.00001	0.153
CERAD Immediate Recall	FCSRT Immediate Recall	<0.00001	0.050
RAVLT Delayed Recall	CERAD Delayed Recall	<0.00001	**0.001**
RAVLT Delayed Recall	FCSRT Delayed Recall	<0.00001	**0.035**
CERAD Delayed Recall	FCSRT Delayed Recall	<0.00001	**0.015**

Figure A.2 Schematic representation of the ROI-based grading method.



AD: Alzheimer's disease, CN: controls, FTD: frontotemporal lobe dementia, DLB: Lewy body dementia.

ROI-based grading method measures the similarity of the ROI (red boxes) from the patient being studied to corresponding ROIs from reference data cases.
The ROI of the patient is represented by a linear combination of ROIs from the reference data. The index is the share of the weights for a certain diagnostics group.

Details about the method can be found from: *Tong T, Wolz R, Coupe P, Hajnal JV, Rueckert D. Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling. Neuroimage. 2013;76:11-23.*

PUBLICATION II

Quantitative evaluation of disease progression in a longitudinal mild cognitive impairment cohort

Runtti H., Mattila J., van Gils M., Koikkalainen J., Soininen H., Lötjönen J. & for the Alzheimer's Disease Neuroimaging Initiative

Journal of Alzheimer's Disease, 2014, 39(1), 49-61
https://doi.org/10.3233/JAD-130359

Journal of Alzheimer's Disease 39 (2014) 49–61
DOI 10.3233/JAD-130359
IOS Press

Quantitative Evaluation of Disease Progression in a Longitudinal Mild Cognitive Impairment Cohort

Hilkka Runtti[a,*], Jussi Mattila[a], Mark van Gils[a], Juha Koikkalainen[a], Hilkka Soininen[b], Jyrki Lötjönen[a] and for the Alzheimer's Disease Neuroimaging Initiative
[a]*VTT Technical Research Centre of Finland, Tampere, Finland*
[b]*Department of Neurology, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland*

Handling Associate Editor: Javier Escudero

Accepted 20 August 2013

Abstract. Several neuropsychological tests and biomarkers of Alzheimer's disease (AD) have been validated and their evolution over time has been explored. In this study, multiple heterogeneous predictors of AD were combined using a supervised learning method called Disease State Index (DSI). The behavior of DSI values over time was examined to study disease progression quantitatively in a mild cognitive impairment (MCI) cohort. The DSI method was applied to longitudinal data from 140 MCI cases that progressed to AD and 149 MCI cases that did not progress to AD during the follow-up. The data included neuropsychological tests, brain volumes from magnetic resonance imaging, cerebrospinal fluid samples, and apolipoprotein E from the Alzheimer's Disease Neuroimaging Initiative database. Linear regression of the longitudinal DSI values (including the DSI value at the point of MCI to AD conversion) was performed for each subject having at least three DSI values available (147 non-converters, 126 converters). Converters had five times higher slopes and almost three times higher intercepts than non-converters. Two subgroups were found in the group of non-converters: one group with stable DSI values over time and another group with clearly increasing DSI values suggesting possible progression to AD in the future. The regression parameters differentiated between the converters and the non-converters with classification accuracy of 76.9% for the slopes and 74.6% for the intercepts. In conclusion, this study demonstrated that quantifying longitudinal patient data using the DSI method provides valid information for follow-up of disease progression and support for decision making.

Keywords: Alzheimer's disease, biomarkers, data mining, decision support techniques, early diagnosis, mild cognitive impairment

INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative disease that develops gradually over the years and finally results in loss of cognitive function and dementia [1]. Mild cognitive impairment (MCI) is an intermediate state between normal cognition and dementia. Patients with MCI have cognitive problems that are not normal for their age and do not yet interfere with their daily activities [2–4]. MCI with memory dysfunction is a risk factor for AD, however, not all MCI patients will progress to AD [2, 3].

There is no cure for AD, but it has been modeled that delaying the onset of the disease would reduce its prevalence considerably, and slowing down its progression would allow more cases to remain as mild AD instead of progressing to moderate or severe AD which

*Correspondence to: Hilkka Runtti, VTT Technical Research Centre of Finland, P.O. Box 1300, FIN-33101 Tampere, Finland. Tel.: +358 40 152 6627; Fax: +358 20 722 3499; E-mail: hilkka.runtti@vtt.fi.

causes huge costs to society [5]. Different treatments to modify disease progression have been studied [6, 7] and it has been shown that they should be started as early as possible to be effective [7, 8]. To make earlier AD diagnosis and interventions feasible, different neuropsychological tests and biomarkers from laboratory tests and imaging have been studied extensively [9–12].

In 2010, Jack et al. [13] proposed a model describing temporal evolution of major AD biomarkers. The model was recently updated on the basis of gained knowledge, and according to it, different biomarkers of AD become abnormal in a certain temporal order and their longitudinal behavior is non-linear [14]. Biomarkers measuring deposition of amyloid-β plaques become abnormal first, years before the clinical symptoms appear. They are followed by indicators of neurodegeneration, and the last biomarkers to become abnormal are structural changes visible in magnetic resonance imaging (MRI) and changes in cerebral metabolism revealed by fluorodeoxyglucose positron emission tomography (FDG-PET). The updated model also takes into account that the severity of cognitive impairment due to pathophysiological load of AD is individual depending on, e.g., genetics, lifestyle, and other brain diseases.

New guidelines, incorporating both cognitive assessment and biomarkers for diagnosing different stages of AD, were recently published as a result of these research findings [15–18]. They state that the detection of preclinical stages of AD in research subjects should be based on biomarkers and that MCI and AD are diagnosed using clinical and cognitive evaluation and biomarkers can provide complementary information.

All the different tests and investigations done in modern diagnostics produce large amounts of data that clinicians need to explore carefully. Assessing the heterogeneous data and measuring longitudinal changes in them may be difficult. Several studies have successfully combined multimodal data to classify subjects into classes of healthy, MCI, or AD using established classification methods, e.g., logistic regression or support vector machines [19–24]. There also exists a statistical Disease State Index (DSI) method which estimates the state of a patient in the continuum from healthy to disease on the basis of measured data. The DSI method has been developed and extensively studied by most of the authors of this manuscript. Mattila et al. [22] demonstrated that it discriminated well between healthy cases, MCI cases that do not convert to AD, MCI cases that convert to AD, and AD cases. A recent study, also by Mattila et al. [25], showed that approximately half of the MCI patients who developed into AD could have been classified with a high accuracy already a year before receiving the clinical diagnoses using the DSI. However, it has not been studied yet how DSI values develop over time in subjects with MCI.

DSI values can be visualized with a Disease State Fingerprint (DSF) technique which shows how results from different tests contribute to the disease state of a patient. The DSF allows rapid interpretation of large amounts of patient data and helps clinicians to discern relevant information from irrelevant [22]. Until now, only data from a single time point have been visualized using the DSF.

The objective of this work was to study disease progression quantitatively using heterogeneous longitudinal data in an MCI cohort. First, it was studied whether it is possible to discern significant trends in the severity of AD as reflected by the DSI and whether subjects that convert from MCI to AD have a different longitudinal DSI behavior than subjects that do not convert. Second, classification of MCI subjects to converters and non-converters on the basis of the trend parameters from longitudinal DSI values was tested. Third, to facilitate interpretation of data, the DSF visualization was developed further for the presentation of longitudinal data.

MATERIALS AND METHODS

Study population

Data used in the analyses were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [26]. ADNI is a 5-year study aiming at developing and testing methods for acquiring and analyzing biological markers that measure the progression of MCI and AD [27]. ADNI was launched in 2004, and approximately 800 subjects of age 50 to 90 years have been recruited at around 50 sites in the United States and Canada. The enrolled subjects included 200 healthy elderly controls, 400 subjects with MCI, and 200 subjects with early AD. The subjects underwent cognitive assessment, neuropsychological testing, and MRI at intervals of six or twelve months for two to four years. Other tests, such as FDG-PET and blood and cerebrospinal fluid samples (CSF), were performed less frequently [28].

In the present study, MCI cases with at least 24 months of follow-up data were included. The selected MCI cases were divided into two groups: a stable

Table 1
Demographics of the study population at the baseline

	Stable MCI	Progressive MCI	*p*
Subjects	149 (51.6%)	140 (48.4%)	
Gender			0.373
Female	51 (34.2%)	55 (39.3%)	
Male	98 (65.8%)	85 (60.7%)	
Age (years)	75.1 ± 7.4	75.4 ± 6.7	0.916
Education (years)	15.9 ± 3.0	15.6 ± 3.0	0.239

Data presented as number of subjects (percentage of subjects %) or mean ± standard deviation. p: Group differences were examined using appropriate tests based on whether their distribution was normal or not as determined by the Kolmogorov-Smirnov test: Pearson χ^2 test (gender) and Mann-Whitney U test (age and education).

MCI group (SMCI, $n = 149$), who did not obtain the diagnosis of AD during the follow-up period, and a progressive MCI group (PMCI, $n = 140$), whose diagnosis changed from MCI to AD during the follow-up. Subjects whose diagnosis changed from MCI to healthy or from MCI to AD and then back to MCI were excluded from the study. Demographics of these two groups are presented in Table 1.

The data were downloaded from the ADNI website (http://adni.loni.ucla.edu) in September 2011. The data used in the analyses comprised Mini-Mental State Examination (MMSE), Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS), Neuropsychological Battery (NeuroBat), brain volume measures based on MRI, amyloid-β and total tau in CSF, and apolipoprotein E (APOE). Details of the included variables are presented in the Supplementary Material. MRI brain volume measures provided to ADNI by Anders Dale Lab (University of California, San Diego) were used. They performed volumetric segmentation of MRI with the FreeSurfer image analysis suite, which is documented and freely available for download online (http://surfer.nmr.mgh.harvard.edu/). Technical details of the segmentation are described in [29].

Diagnosis of MCI and AD in the ADNI is based on evaluation of memory, cognition, and functional performance (memory complaints by a subject or a study partner, Logical Memory II, MMSE, and Clinical Dementia Rating) [28]. In addition, diagnosis of probable AD requires fulfillment of the AD criteria defined by the NINCDS-ADRDA (the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association) [30, 31]. Although the diagnosis is partly based on MMSE and Logical Memory II, they were included in the data analyses in this study because 1) MMSE is widely used making it interesting in clinical sense, 2) the diagnosis is not based only on the MMSE and Logical Memory II, and 3) the ADNI criteria to decide between MCI versus AD does allow overlap in MMSE score and Logical Memory II score.

Variables summarizing the tests, e.g., total MMSE score and ADAS 13 point total, were excluded as independent variables from the analysis because the subscores and the individual items contain the same information as the total scores. Justification for the use of individual items instead of total scores is that some items may differentiate between SMCI and PMCI cases better than others and part of the available information is lost if only the total scores are used. For example, Llano et al. [32] weighted individual items of ADAS with coefficients derived using data-driven approach and constructed a new composite ADAS score. Their composite score differentiated normal controls, MCI, and AD cases better than the ADAS total score and the composite score also predicted conversion to AD slightly better than the ADAS total score.

Disease State Index

The DSI is a statistical method for deriving a scalar value that estimates the state of a disease in a patient [22]. The DSI method is based on the computation of two different values: DSI values and relevance values. The DSI value of an individual variable is computed by comparing a measurement value from a patient to the distributions of known healthy and diseased cases using a so-called fitness function. DSI values are between zero and one, with higher values indicating that the patient fits better to the disease than to the control population on the basis of the measured data. The relevance value describes how well the variable differentiates between the known healthy and diseased cases. In other words, relevance is a measure of the differences in the data measured from healthy and diseased cases. Relevance values, like the DSI values, are also between zero and one, with higher values representing better discrimination. A composite DSI combining different variables is computed as a weighted arithmetic mean of the individual DSI values weighted by the relevance values. This averaging is done several times recursively to yield a hierarchy of DSI values that reveals the overall position or rank in relation to the disease, i.e., quantifies the progression of a disease based on available patient data. In this work, the study population consisted of SMCIs as control cases and PMCIs as disease cases.

The DSI method is robust against overfitting by its design. Estimation of the DSI and relevance values

Table 2
Number of available patient visits at different time points

	Baseline	Month 6	Month 12	Month 18	Month 24	Month 30	Month 36	Month 42	Month 48
Total	289	287	287	279	281	0	233	0	51
SMCI	149	148	147	143	142	0	121	0	19
PMCI	140	139	140	136	139	0	112	0	32

SMCI, stable mild cognitive impairment; PMCI, progressive mild cognitive impairment.

for individual variables is done independently from other variables, thus, there is no over-dimensionality at the variable level because only two parameters are estimated for each variable (the DSI value and the relevance value). In addition, weighting of features and the use of the hierarchy lead in practice to feature selection. As a result, any few values alone will not determine the resulting composite DSI value, but it is an amalgam of all relevant data sources. Mathematical details of the computation of the DSI and relevance values are explained in [22].

The DSI values can be calculated on the basis of a single variable or multiple variables together. In this study, it was investigated whether combining different data modalities would yield better results than utilizing data from a single modality alone. Thus, DSI values were calculated using two different approaches: 1) using all available variables together (MMSE, ADAS, NeuroBat, MRI, CSF, and APOE) and 2) using data from individual data modalities independently (MMSE, ADAS, NeuroBat, and MRI). CSF was measured less frequently so it was not analyzed individually and neither was APOE genetics, which do not change with disease progression. For the calculation of the DSI values, subjects were divided into ten training and test sets for stratified 10-fold cross-validation in which each fold contains the same proportions of class labels. The training data used for building the model of AD progression included actual measurement values from SMCI baseline visits and actual measurement values from the time of receiving AD diagnosis for PMCI cases. This kind of selection of training data sets the dynamic range of the DSI method between SMCIs at the baseline and early AD, i.e., the dynamic range of the DSI method was optimized for the purposes of the study and clinical problem at the hand. The test sets included data from the complete series of visits of the remaining SMCI and PMCI cases. The number of patient visits available at the different time points is shown in Table 2. Missing values in the raw data (e.g., a missing result in MMSE) were replaced with the values from the patient's previous available visit. This allowed having complete data sets for the analysis at each patient visit. Although using previous data can result in slightly outdated data and conservative disease progression estimates for some patient visits, that data were known to have been available at those time points.

Disease State Fingerprint

The DSF is a method for visualizing the patient data and the hierarchy of the DSI values [22]. Example visualizations are shown in the left panel of Fig. 1. DSF consists of a tree with nodes of different sizes and colors. The size of the node indicates the relevance value, i.e., how well a variable or a test differentiates between SMCI and PMCI, and color indicates the DSI value. Higher DSI values refer to PMCI and result in shades of red. Lower values represent SMCI and result in shades of blue. In this study, the progression of AD was visualized using the DSF technique extended with support for longitudinal data.

Synchronization of the time stamps

The initial visits of MCI patients to a memory clinic occurred in different phases of the disease. For example, some PMCI cases converted from MCI to AD at follow-up month 6 and others at month 36. To take this into account, the time stamps of the patient visits were synchronized. The moment of receiving AD diagnosis was set as the zero time point (Z) of PMCIs. For SMCIs, the last available time point up to month 36 was set as their Z. The time points preceding the zero point were labeled as Z-6, Z-12, etc. DSI values from Z-42 and Z-48 months were excluded from the analysis because they contained only a few cases. Thus, DSI values computed from visit data at Z, Z-6, Z-12, Z-18, Z-24, Z-30, and Z-36 months were used in the analysis. Only those subjects who had at least three DSI values available in all approaches (DSI calculated using all variables, MMSE, ADAS, NeuroBat, or MRI), were included for further analysis. The purpose was to perform linear regression (see below) and using only two points would have yielded in perfect regression, making the comparison of goodness of fit values between the different datasets unfair. The number of available

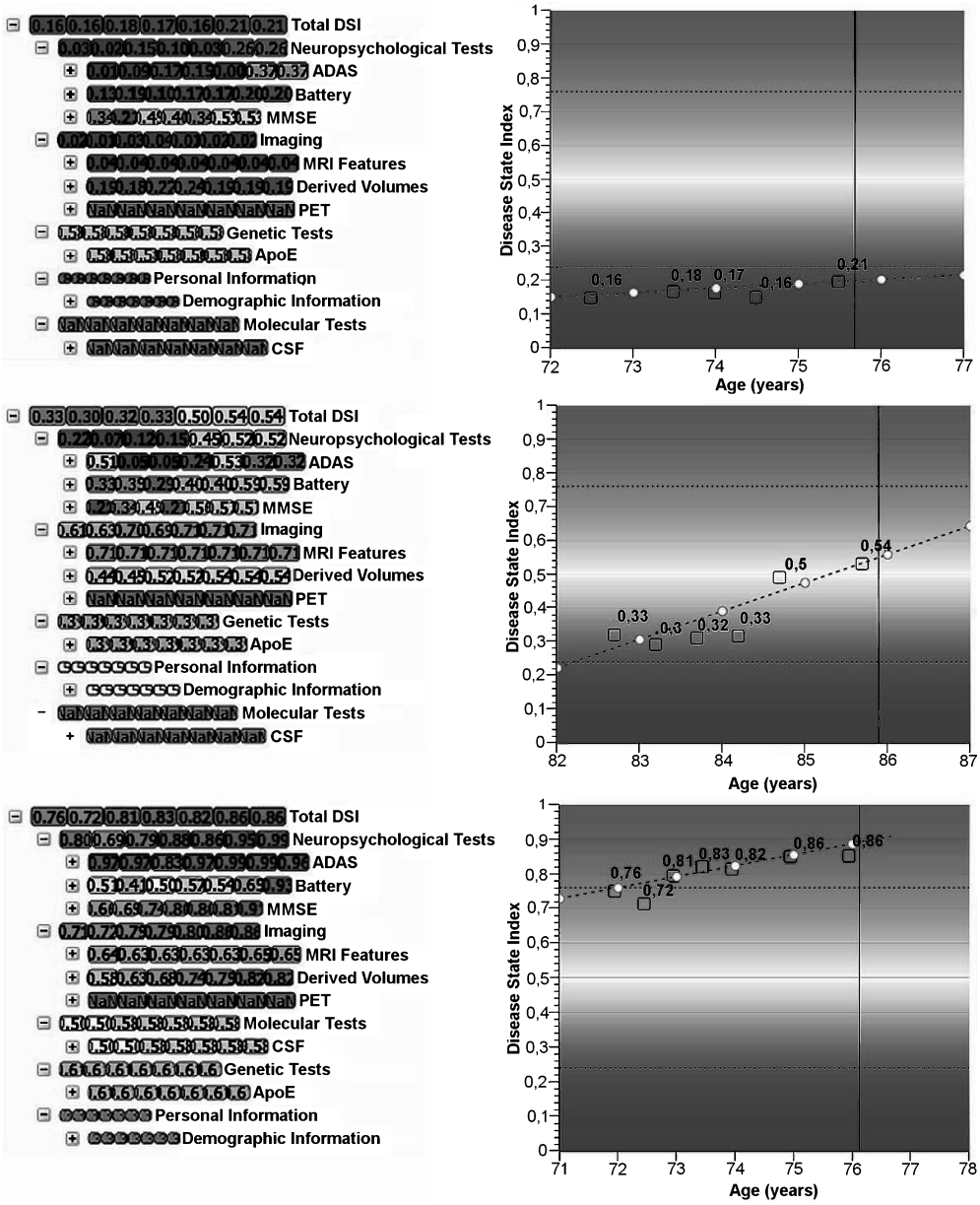


Fig. 1. Visualizations of three sets of longitudinal patient data. Left panel: Disease State Fingerprints (DSF) in which Disease State Index (DSI) values of the individual tests at different time points are shown on the rows. Total DSI values (the topmost rows of the DSFs) combines the results from the individual tests. Sizes of the boxes indicate how well the variable discriminates between the stable (SMCI) and progressive (PMCI) mild cognitive impairment cases. Color indicates to which group the data fits the best. Blue corresponds to SMCI and red to PMCI. Right panel: linear regression of the total DSI values (red dashed line with white circles). Black squares present the total DSI values of a patient. The horizontal lines indicate a threshold where the classification accuracy of 85% is achieved. The vertical line shows the current age of a patient. Data from two SMCI cases are presented in the topmost panels and data from a PMCI case is presented in the lowest panel.

Table 3
Number of Disease State Index values of the SMCI and PMCI cases at synchronized time points. The last available time point up to month 36 was selected as the zero time point (Z) of SMCI. The moment of receiving Alzheimer's disease diagnosis was set as the Z of PMCI. The time points preceding the Z were labeled as Z-6 etc

	Z-36	Z-30	Z-24	Z-18	Z-12	Z-6	Z
SMCI	147	147	147	147	147	147	147
PMCI	29	29	64	90	126	126	126

SMCI, stable mild cognitive impairment; PMCI, progressive mild cognitive impairment. The number of SMCI cases stays the same because the visit Z-36 is their baseline visit and any missing values have been replaced with the values from the previous available visit. The number of PMCI cases changes over time because some have converted in an early phase of the study. Only the cases having at least three available DSI values were included.

DSI values of the included SMCI and PMCI cases at the synchronized time points is presented in Table 3.

Modeling progression of AD

In this work, it was assumed that the change of the DSI values over time, and thus the progression of AD, can be modeled linearly:

$$DSI = a * t + b \quad (1)$$

where a is the slope of regression (rate of change for DSI values), b is the intercept of regression (DSI value at the time point zero), and t is time measured in months. A linear model was selected because it is the simplest method to model the progression of AD and it is also the simplest to interpret. Another reason was that due to the synchronization of the time stamps some subjects had only few DSI values available for the regression. Thus, there were not enough data points for more complicated models. The third reason supporting the linear model was that the follow-up times were relatively short compared with the time span of disease progression in AD in overall. Linear regression was performed for each subject separately to model each individual's disease progression.

Differentiation using the trend parameters

Classification of subjects as SMCI or PMCI cases on the basis of their regression parameters (slope, intercept) was studied as follows. First, optimal classification thresholds for the regression parameters were defined on the basis of the receiver operating characteristic (ROC) curves. Then, the regression parameters were compared to the threshold value and if it was exceeded the subject was classified as PMCI. Otherwise he or she was classified as SMCI. The thresholds and classification performance measures (classification accuracy, sensitivity, and specificity) were calculated using the stratified 10-fold cross-validation.

Statistical methods

Normality of the continuous demographic variables was studied using Kolmogorov-Smirnov test. Group differences in demographics between SMCI and PMCI groups were examined using non-parametric Mann-Whitney U test for continuous variables and Pearson χ^2 test for categorical variables.

Linear regression was performed using the longitudinal DSI values which were derived using 1) all available variables together (total) and 2) data from individual tests separately. Goodness of fit of the linear regression using 1) and 2) was compared using R^2, adjusted R^2, and mean square errors. Residuals of the regression were also examined using histograms and by plotting residuals versus predicted values. The regression parameters of the SMCI and PMCI groups were compared to zero using one-sample Wilcoxon Signed Rank test and the differences between the groups were studied using Mann-Whitney U test.

Normality of the regression parameters was studied using histograms. On the basis of the initial histogram analysis, it appeared that the slopes of the SMCI group may have a bimodal distribution. Fits of unimodal and bimodal distributions were compared and details of these analyses are explained in the Supplementary Material.

Subjects were classified as SMCIs or PMCIs on the basis of their regression parameters. Classification performance was measured using the area under the ROC curve (AUC), classification accuracies, sensitivities, and specificities. To study whether using all data modalities together would yield in significantly greater classification performance than using only a single data modality, classification accuracies of the individual tests were compared to the classification accuracies derived using all data. Thus, four comparisons with both the slopes and the intercepts (total-MMSE, total-ADAS, total-NeuroBat, total-MRI) were performed. The classification accuracies of the slopes and the intercepts derived using all data were also compared. Paired samples t-test was used if the classification accuracies were normally distributed according to Kolmogorov-Smirnov test, otherwise, related-samples Wilcoxon Signed Rank test was performed. In all analyses, $p < 0.05$ was considered significant. In pairwise comparisons of classification accuracies, Bonferroni

Table 4
Goodness of fit for the linear regression of longitudinal Disease State Index values derived using different data modalities

Dataset	R^2	Adjusted R^2	Mean square error
Total	0.553 ± 0.289	0.422 ± 0.369	0.006 ± 0.008
MMSE	0.364 ± 0.295	0.172 ± 0.390	0.014 ± 0.016
ADAS	0.388 ± 0.298	0.196 ± 0.413	0.024 ± 0.026
NeuroBat	0.475 ± 0.318	0.315 ± 0.426	0.005 ± 0.004
MRI	0.721 ± 0.259	0.642 ± 0.321	0.001 ± 0.001

Total, All available variables included when calculating DSI values; MMSE, Mini-Mental State Examination; ADAS, Alzheimer's Disease Assessment Scale-cognitive subscale; NeuroBat, Neuropsychological Battery; MRI, brain volumes derived from magnetic resonance imaging. The values are mean ± standard deviation because the linear regression was performed for each subject independently.

Table 5
Regression parameters of longitudinal Disease State Index values for SMCI and PMCI groups

	SMCI	PMCI
Slope*	0.002 (0.000, 0.006)$^+$	0.010 (0.005, 0.015)$^+$
Intercept*	0.295 (0.139, 0.621)$^+$	0.754 (0.626, 0.860)$^+$
n	7 (7; 7)	5 (3; 5)

Values are median (25th percentile, 75th percentile). SMCI, stable mild cognitive impairment; PMCI, progressive mild cognitive impairment, n, number of points in the regression, *statistically significant difference between the groups (Mann-Whitney U test, $p < 0.0005$), $^+$significantly different from zero (one-sample Wilcoxon Signed Rank test, $p < 0.0005$). Disease State Index values were derived using all variables together.

correction was applied and $p < 0.0056$ was considered significant (number of comparisons was nine).

All analyses were performed in Matlab R2012a (The Mathworks, Natick, MA) and IBM SPSS Statistics 19 (IBM, Armonk, NY). Visualizations were processed in GNU Image Manipulation Program 2.0 (GIMP 2.0, freely available at http://www.gimp.org/).

RESULTS

Modeling progression of AD

Goodness of fit for linear regression of the longitudinal DSI values is shown in Table 4. On the basis of R^2, adjusted R^2, and mean square error, the linear association was the strongest when DSI values were calculated using only MRI-derived volumes. The linear model fitted the second best when all available variables were used together. The longitudinal DSI values derived on the basis of cognitive and neuropsychological tests had the smallest association values. Plots of residuals versus predicted values supported the interpretation that the DSI values calculated on the basis of ADAS and MMSE were the least linear over time: points in the plots were not as randomly distributed as they were when the DSI values were based on all available data, MRI, or NeuroBat (results not shown here).

The linear regression of the DSI values over time was performed for each subject independently. Medians of the regression parameters for SMCI and PMCI groups are shown in Table 5. The slopes and the intercepts of both groups were higher than zero ($p < 0.0005$). There were also clear differences between the two groups: PMCIs had five times higher slopes and almost three times higher intercepts than SMCIs ($p < 0.0005$).

The distributions of the slopes of both groups are presented in Fig. 2. On the basis of the visual inspection, the SMCI curve deviated from a Gaussian distribution containing also cases with higher slopes. Therefore, a hypothesis was put forth that the SMCI group actually contained two subgroups: one with truly stable DSI values and one with non-stable DSI values having signs of disease progression. A mixture distribution of two normal curves was fitted to the slopes of the SMCIs. The fits of unimodal and bimodal distributions were compared, and the results and estimated parameters are shown in the Supplementary Material. The results showed that the bimodal distribution fitted better to the slopes of the SMCIs than the unimodal distribution supporting the idea that two subgroups do exist within the SMCI group.

Visualizing progression of AD

In Fig. 1, the progression of AD is visualized using the DSF and the regression line of the DSI values. Most of the nodes in the DSF of a clear SMCI case are blue indicating that the patient data remained constantly unlike the data of those with AD. Also, the slope and the intercept of the regression line have low values (Fig. 1, topmost panel). On the contrary, almost all nodes of a clear PMCI case are red, indicating strong resemblance to previously diagnosed AD cases, and the slope and the intercept are higher as well (Fig. 1, lowest panel). A SMCI case with clearly increasing DSI values and the DSF changing from blue to red is also shown (Fig. 3, mid-panel). This case belongs to the subgroup of SMCI cases with non-stable DSI values in Fig. 2.

Differentiation using the trend parameters

MCI cases were classified as SMCI or PMCI using the regression parameters of the longitudinal DSI values, and the classification performance results are

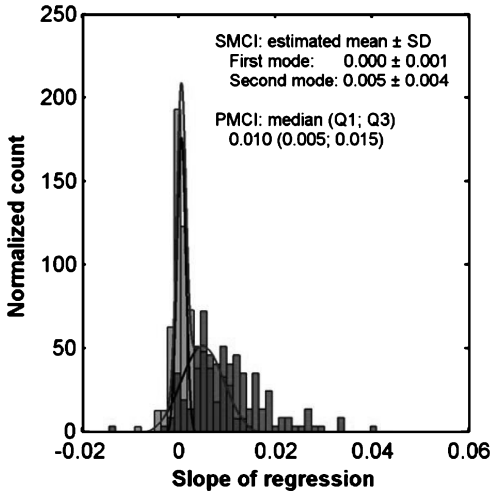


Fig. 2. Histograms of the slopes for stable (SMCI, blue) and progressive (PMCI, red) mild cognitive impairment cases. There appears to be two separate subgroups in the SMCI group. A mixture distribution of two normal curves fitted to the slopes of SMCIs is also shown. The areas of the histograms are scaled to one. (SD = standard deviation, Q1 = 25th quartile, Q3 = 75th quartile).

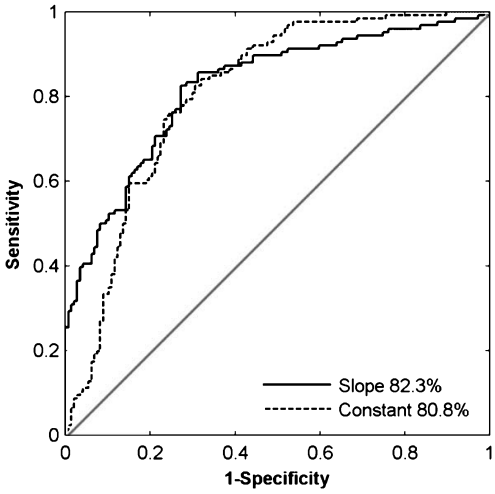


Fig. 3. Receiver operating characteristic curves of the slope (solid line) and the intercept (dashed line). Regression parameters were defined using total Disease State Index values over time.

presented in Table 6. AUCs were the highest when all available variables were used in the analysis (total). Classification accuracies were normally distributed, except for the slopes derived using NeuroBat. The

Table 6
Classification performance of the regression parameters of the longitudinal Disease State Index values derived using different datasets

	AUC (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)
Slope				
Total	82.3	76.9 ± 8.8	82.2 ± 13.7	73.0 ± 15.0
MMSE	77.1	71.8 ± 7.6	55.5 ± 15.5	86.5 ± 5.5
ADAS	76.8	68.7 ± 10.2	51.1 ± 19.2	83.6 ± 10.2
NeuroBat	76.6	69.2 ± 5.8	60.2 ± 13.2	76.9 ± 15.3
MRI	71.0	66.8 ± 8.1	49.5 ± 14.4	80.6 ± 14.7
Intercept				
Total	80.8	74.6 ± 8.7	75.1 ± 17.4	74.4 ± 12.2
MMSE	79.0	72.0 ± 5.0	84.2 ± 11.6	61.5 ± 11.6
ADAS	80.3	74.9 ± 8.8	74.4 ± 15.6	75.7 ± 10.5
NeuroBat	79.3	66.9 ± 6.1	74.4 ± 21.7	61.0 ± 14.0
MRI	69.6	60.4 ± 8.9	55.6 ± 16.2	63.9 ± 16.2

Results are mean ± standard deviation from the stratified 10-fold cross-validation, except for the AUC. Total, all available variables included when calculating Disease State Index values; MMSE, Mini-Mental State Examination; ADAS, Alzheimer's Disease Assessment Scale-cognitive subscale; NeuroBat, Neuropsychological Battery; MRI, brain volumes derived from magnetic resonance imaging; AUC, area under the receiver operating characteristic curve.

classification accuracy of the slopes (total) was significantly higher than the classification accuracies of the slopes derived using ADAS or MRI ($p = 0.001$ for total-ADAS and $p = 0.005$ for total-MRI comparisons). The classification accuracy of the intercepts (total) was significantly higher than classification accuracy of the MRI-derived intercepts ($p = 0.004$). Other pairwise comparisons of the slopes and the intercepts were not statistically significant (all $p > 0.01$, Bonferroni-corrected significance level was 0.0056). The classification accuracies of the slopes (total) and the intercepts (total) were very similar (76.9% and 74.6%, respectively, $p = 0.309$). ROC curves of the slopes (total) and the intercepts (total) are presented in Fig. 3.

DISCUSSION

Quantification of disease progression from MCI to AD was studied by applying the DSI method to heterogeneous longitudinal patient data and analyzing the behavior of the DSI values over time in subjects with MCI. Trend parameters of the longitudinal DSI values were obtained from regression and ability of them to differentiate between the groups of stable and progressive MCI was also studied.

In this study, it was assumed that the behavior of the longitudinal DSI values can be modeled linearly. The linear association was the strongest when the DSI values were based only on MRI features. Behavior of the total DSI values was not as linear because

neuropsychological tests were included and their temporal behavior was the least linear. The linear model may not necessarily be the best model for progression of AD but it was selected because of simplicity and due to paucity of data. Some subjects with PMCI had only a few DSI values available for the regression due to synchronization of the time stamps.

Jack and his colleagues [13] proposed that changes in biomarkers over time would be sigmoidal and biomarkers would become abnormal in a certain temporal order. These assumptions gained support in several studies and they still are core components of the recently revised model [14]. Caroli et al. [33] provided the first evidence supporting the first version of the model. They compared the fit of linear and sigmoidal model and concluded that the sigmoidal model fitted better for hippocampal volume, and amyloid-β and total-tau in CSF. The linear model fitted better for FDG-PET data. Instead of real longitudinal data, Caroli et al. [33] used data from healthy controls, PMCIs, and early and late ADs at the baseline to reflect the progression of AD. Mouiha and Duchesne [34] used the same kind of cross-sectional setting to study the relationship between biomarkers and disease severity. They fitted six different models (linear, quadratic, robust quadratic, local quadratic regression, penalized B-spline, and sigmoid) to baseline data from healthy controls, PMCI, and AD cases [34]. According to them, amyloid-β had a piece-wise quadratic relationship, hippocampal volume and CSF measures of phosphorylated tau and total tau were best modeled with penalized B-splines, and linear model was the best fit for FDG-PET [34].

The results in this study show that the change of DSI values over time as reflected by the slope of the linear regression equation is clearly different in the SMCI and PMCI groups. The slope of PMCI cases was five times higher than the slope of SMCI cases. When the slopes of SMCI cases were studied more thoroughly, it was noticed that there were two different subgroups in the SMCI group: a group with lower slopes and another group with higher slopes that overlap with the slopes of the PMCI cases. It is expected that the peak with higher slopes represents MCIs that would convert to AD or other dementia later if the follow-up was continued. Davatzikos et al. [20] and Cui et al. [19] also found in their studies that subjects in the SMCI group did not have uniform results. Some SMCI cases had markers similar to AD, suggesting that they may convert to AD in the future [19, 20].

Samtani et al. [35] modeled a subject's rate of disease progression using a logistic model with several covariates. Severity of the disease was measured using ADAS and the analysis was restricted to an AD population [35]. Another approach for modeling disease progression was presented by Escudero et al. [36]. They found profiles of disease and normality using an unsupervised learning method (k-means clustering). Escudero et al. [36] calculated a so-called Bioindex that describes a subject's degree of membership to the profile of disease on the basis of measured data. To study evolution of Bioindeces over time, a sigmoid function was fitted to the Bioindex values at different time points. They used the same approach as here and fitted an individual function to the Bioindeces of each subject and studied evolution of Bioindeces in the groups of SMCI and PMCI. As in this study, they found that converters had steeper progression towards AD than non-converters. However, Escudero et al. [36] did not take into account that MCI patients arrived in the study at different phases of the disease, and they did not synchronize the time stamps as we did.

Patient visits in this study were synchronized according to the time of receiving AD diagnosis. Using this method, the accuracy of the synchronization depends on the accuracy of the actual AD diagnoses. Also, data points of the SMCI cases are not synchronized because they do not have an AD diagnosis. Jedynak et al. [37] and Yang et al. [38] proposed more sophisticated methods for synchronization. Jedynak et al. [37] used multiple biomarkers to create a disease progression score, which set the subjects on the same timeline [37]. Biomarkers were assumed to follow a sigmoidal function when constructing the disease progression score [37]. Yang et al. [38] modeled evolution of ADAS 13 score over time with an exponential model and then defined the start of the cognitive decline using the model. Other biomarkers were then synchronized using the estimated period of cognitive decline. After the synchronization, evolution of biomarkers over time and relations between them were clearer and they supported the model presented by Jack et al. [13, 14, 38]. In the approach presented in [38], one needs to define an accurate model for the progression of ADAS 13 score over time, and the accuracy of the synchronization depends on the suitability of the model.

The dynamic range for the DSI depends on training sets used. In this study, the DSI values were calculated on the basis of data from SMCI cases at baseline and PMCI cases at the point of conversion to AD. Thus, the dynamic range lies between MCI and early AD. Using the same model of disease progression to study healthy controls and late AD groups would saturate DSI values close to zero and one, respectively. On the

other hand, if the training set consisted of PMCI and AD groups, the DSI would characterize changes at the later phase of the disease. Thus, if different training sets are used, the longitudinal behavior of the DSI values can be somewhat different. As another example, if training set included healthy and AD cases, slopes of the SMCI and PMCI groups should be closer to each other than they are in this study.

Training data for this study was selected from SMCI cases at the baseline and PMCI cases at the point of conversion because the initial purpose for the proposed method is in early diagnosis of AD. The main use case for the method is a situation where a subject with memory complaints arrives at a clinic. After some tests have been administered, computer-based decision support tools could help in objective assessment of patient data and possibly provide help for earlier diagnosis of AD. If the diagnosis cannot be made at the baseline, longitudinal quantification of progressing disease state provides additional information to base the diagnosis on. By selecting SMCI cases at the baseline and PMCI cases at the moment of receiving diagnosis as the training set, the system is optimized to detect early AD cases from an MCI population referred to a memory clinic. The DSI method is currently incorporated in a decision support tool that will be used in pilot studies and the training set used in the tool comprises SMCI and PMCI cases, similar to this study. When studies with other purposes (e.g., focus on conversion from normal cognition to MCI) are done in the future, then the practical issues of selecting the most appropriate training population will be addressed.

Recently, several studies have predicted the conversion from MCI to AD by combining multiple data modalities and identifying converters and non-converters on the basis of the data [19–23]. In these studies, multimodal data were combined using logistic regression [21, 22], the DSI method [22], support vector machine classifiers [19, 22, 23, 39], and a Naive Bayes classifier [22]. In [19, 20, 22, 40], it was found that combination of multimodal data resulted in better classification performance than the use of a single modality of data, e.g., using only neuropsychological tests. However, those studies did not report whether the differences were statistically significant. Ewers et al. [21] found that increasing number of variables in the model from one to four increased the classification accuracy, but the increase was not significant according to the 95% confidence intervals. Cui et al. [39] also combined different data modalities for predicting conversion from normal cognition to MCI. They reported that combination of neuropsychological test scores and MRI features resulted in significantly higher classification accuracy for the predictions than using either of the data modalities alone. Results from our study are in line with the previous research findings. Combination of all available data resulted in higher classification accuracies and AUCs than using only a single modality of data and increases in classification accuracies were not always statistically significant. To account for multiple comparisons, we used Bonferroni correction which is known to be a rather conservative method. However, in many comparisons, p-values were higher than 0.05.

It is worth noting that the calculation of the linear regression included DSI values from the point of conversion for PMCI cases. Thus, the classification performance measures presented here do not describe the ability of the trend parameters to predict conversion from MCI to AD. However, they demonstrate that the trend parameters of the DSI values are clearly different between the groups of SMCI and PMCI. Prediction of MCI to AD conversion with the DSI method using data from the ADNI database has already been studied in [22] and [25].

One interesting finding was that the MRI-derived longitudinal DSI values had the strongest linear association but the regression parameters of the MRI-based DSI values performed the worst in the classification. One explanation could be that changes related to normal aging in the brain may interfere with the results. For example, Koikkalainen et al. [41] removed effects of age and other confounding factors by dividing patients into subgroups and using linear regression. These procedures improved classification accuracies in their study. Another explanation could be that MRI may be a better indicator of the rate of disease progression than of the disease stage. Stronger linearity of the MRI-derived DSI over time might also be caused by the fact that MRI measures are not as prone to daily variations as neuropsychological tests may be.

Missing values were imputed with the values from the previous available visit. This approach resulted in slightly outdated data for some patient visits and biased the results towards non-progression. This approach was chosen so that all data used in the analyses really were available from a patient at the specific moments. This would not be the case, e.g., if missing values were replaced with the next available values or using other more complex imputation methods. Replacing missing values with next available values would have biased results toward progression to some extent and there would still have been missing values because some patients did not have any values available beyond the

last time point. If the missing values had not been imputed at all, the DSI values at different time points would have been calculated using different variables for each visit and this would have hindered the interpretation of the longitudinal results.

The study had some limitations. The final diagnoses for the subjects were determined on the basis of clinical evaluation and they were not verified with postmortem histological samples taken from the brain. Also, the study period of 48 months is relatively short. Thus, some subjects diagnosed currently as stable MCI may convert to AD later. This study utilized longitudinal data from a period of 2–4 years. In clinics, where the patients are diagnosed, there may not be data from such a long period available. Less longitudinal data will probably produce more variation in the slopes and the intercepts of the regression equation. On the other hand, this study suggests that quantifying longitudinal patient data using the DSI method provides valid information for decision support and is a valid methodology to follow-up a patient's condition in a quantitative manner.

In conclusion, this study demonstrates that combining sparse and heterogeneous data with the DSI method can be used for deriving a quantitative measure related to early AD progression. Significant trends were found in longitudinal DSI values: rate of change of DSI values was five times higher in the PMCI group than in the SMCI group. Classification of the subjects as converters and non-converters on the basis of the regression parameters (the slope and the intercept) also showed that SMCI and PMCI cases can be differentiated on the basis of the trend parameters.

ACKNOWLEDGMENTS

This work was partially funded under the 7th Framework Programme by the European Commission (http://cordis.europa.eu/ist; EU-Grant-224328-PredictAD; Name: From Patient Data to Personalized Healthcare in Alzheimer's Disease) and the SalWe Research Program for Mind and Body (Tekes – The Finnish Funding Agency for Technology and Innovation, grant 1104/10).

Authors' disclosures available online (http://www.j-alz.com/disclosures/view.php?id=1923).

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Alzheimer's Association; Alzheimer's Drug Discovery Foundation; BioClinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Synarc Inc.; and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (http://www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129 and K01 AG030514.

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

SUPPLEMENTARY MATERIAL

The supplementary material and tables are available in the electronic version of this article: http://dx.doi.org/10.3233/JAD-130359.

REFERENCES

[1] Nestor PJ, Scheltens P, Hodges JR (2004) Advances in the early detection of Alzheimer's disease. *Nat Med* **10** Suppl, S34-S41.

[2] Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, Belleville S, Brodaty H, Bennett D, Chertkow H, Cummings JL, de Leon M, Feldman H, Ganguli M, Hampel H, Scheltens P, Tierney MC, Whitehouse P, Winblad B, International Psychogeriatric Association Expert Conference

on mild cognitive impairment (2006) Mild cognitive impairment. *Lancet* **367**, 1262-1270.
[3] Petersen RC (2009) Early diagnosis of Alzheimer's disease: is MCI too late? *Curr Alzheimer Res* **6**, 324-330.
[4] Brooks LG, Loewenstein DA (2010) Assessing the progression of mild cognitive impairment to Alzheimer's disease: current trends and future directions. *Alzheimers Res Ther* **2**, 28.
[5] Sloane PD, Zimmerman S, Suchindran C, Reed P, Wang L, Boustani M, Sudha S (2002) The public health impact of Alzheimer's disease, 2000-2050: potential implication of treatment advances. *Annu Rev Public Health* **23**, 213-231.
[6] Salloway S, Mintzer J, Weiner MF, Cummings JL (2008) Disease-modifying therapies in Alzheimer's disease. *Alzheimers Dement* **4**, 65-79.
[7] Galimberti D, Scarpini E (2011) Disease-modifying treatments for Alzheimer's disease. *Ther Adv Neurol Disord* **4**, 203-216.
[8] Duara R, Barker W, Loewenstein D, Bain L (2009) The basis for disease-modifying treatments for Alzheimer's disease: the Sixth Annual Mild Cognitive Impairment Symposium. *Alzheimers Dement* **5**, 66-74.
[9] Ballard C, Gauthier S, Corbett A, Brayne C, Aarsland D, Jones E (2011) Alzheimer's disease. *Lancet* **377**, 1019-1031.
[10] Hampel H, Bürger K, Teipel SJ, Bokde AL, Zetterberg H, Blennow K (2008) Core candidate neurochemical and imaging biomarkers of Alzheimer's disease. *Alzheimers Dement* **4**, 38-48.
[11] Borroni B, Premi E, Di Luca M, Padovani A (2007) Combined biomarkers for early Alzheimer disease diagnosis. *Curr Med Chem* **14**, 1171-1178.
[12] Craig-Schapiro R, Fagan AM, Holtzman DM (2009) Biomarkers of Alzheimer's disease. *Neurobiol Dis* **35**, 128-140.
[13] Jack CR Jr, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, Petersen RC, Trojanowski JQ (2010) Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol* **9**, 119-128.
[14] Jack CR Jr, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, Shaw LM, Vemuri P, Wiste HJ, Weigand SD, Lesnick TG, Pankratz VS, Donohue MC, Trojanowski JQ (2013) Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol* **12**, 207-216.
[15] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo T, Jack CR Jr, Kaye J, Montine TJ, Park DC, Reiman EM, Rowe CC, Siemers E, Stern Y, Yaffe K, Carrillo MC, Thies B, Morrison-Bogorad M, Wagster MV, Phelps CH (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. *Alzheimers Dement* **7**, 280-292.
[16] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC, Snyder PJ, Carrillo MC, Thies B, Phelps CH (2011) The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 270-279.
[17] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, Weintraub S, Phelps CH (2011) The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 263-269.
[18] Jack CR Jr, Albert MS, Knopman DS, McKhann GM, Sperling RA, Carrillo MC, Thies B, Phelps CH (2011) Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 257-262.
[19] Cui Y, Liu B, Luo S, Zhen X, Fan M, Liu T, Zhu W, Park M, Jiang T, Jin JS, the Alzheimer's Disease Neuroimaging Initiative (2011) Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors. *PLoS One* **6**, e21896.
[20] Davatzikos C, Bhatt P, Shaw LM, Batmanghelich KN, Trojanowski JQ (2011) Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification. *Neurobiol Aging* **32**, 2322.e19-2322.e27.
[21] Ewers M, Walsh C, Trojanowski JQ, Shaw LM, Petersen RC, Jack CR Jr, Feldman HH, Bokde AL, Alexander GE, Scheltens P, Vellas B, Dubois B, Weiner M, Hampel H, North American Alzheimer's Disease Neuroimaging Initiative (ADNI) (2012) Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiol Aging* **33**, 1203-1214.
[22] Mattila J, Koikkalainen J, Virkki A, Simonsen A, van Gils M, Waldemar G, Soininen H, Lötjönen J, the Alzheimer's Disease Neuroimaging Initiative (2011) A disease state fingerprint for evaluation of Alzheimer's disease. *J Alzheimers Dis* **27**, 163-176.
[23] Zhang D, Shen D, the Alzheimer's Disease Neuroimaging Initiative (2012) Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. *PLoS One* **7**, e33182.
[24] Gray KR, Aljabar P, Heckemann RA, Hammers A, Rueckert D, the Alzheimer's Disease Neuroimaging Initiative (2012) Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *Neuroimage* **65**, 167-175.
[25] Mattila J, Soininen H, Koikkalainen J, Rueckert D, Wolz R, Waldemar G, Lötjönen J, the Alzheimer's Disease Neuroimaging Initiative (2012) Optimizing the diagnosis of early Alzheimer's disease in mild cognitive impairment subjects. *J Alzheimers Dis* **32**, 969-979.
[26] Alzheimer's Disease Neuroimaging Initiative, http://adni.loni.ucla.edu, Accessed on May 4, 2013.
[27] Weiner MW, Aisen PS, Jack CR Jr, Jagust WJ, Trojanowski JQ, Shaw L, Saykin AJ, Morris JC, Cairns N, Beckett LA, Toga A, Green R, Walter S, Soares H, Snyder P, Siemers E, Potter W, Cole PE, Schmidt M, the Alzheimer's Disease Neuroimaging Initiative (2010) The Alzheimer's Disease Neuroimaging Initiative: progress report and future plans. *Alzheimers Dement* **6**, 202-211.
[28] ADNI1 Procedures Manual (2011), http://www.adni-info.org/Scientists/Pdfs/ADNI1_Procedures_Manual_Revised_12052011.pdf, Posted on May 12, 2011, Accessed on January 23, 2012.
[29] Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM (2002) whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341-355.

[30] Dubois B, Feldman HH, Jacova C, DeKosky ST, Barberger-Gateau P, Cummings J, Delacourte A, Galasko D, Gauthier S, Jicha G, Meguro K, O'Brien J, Pasquier F, Robert P, Rossor M, Salloway S, Stern Y, Visser PJ, Scheltens P (2007) Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. *Lancet Neurol* **6**, 734-746.

[31] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984) Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**, 939-944.

[32] Llano DA, Laforet G, Devanarayan V, the Alzheimer's Disease Neuroimaging Initiative (2011) Derivation of a new ADAS-cog composite using tree-based multivariate analysis: prediction of conversion from mild cognitive impairment to Alzheimer disease. *Alzheimer Dis Assoc Disord* **25**, 73-84.

[33] Caroli A, Frisoni GB, the Alzheimer's Disease Neuroimaging Initiative (2010) The dynamics of Alzheimer's disease biomarkers in the Alzheimer's Disease Neuroimaging Initiative cohort. *Neurobiol Aging* **31**, 1263-1274.

[34] Mouiha A, Duchesne S, the Alzheimer's Disease Neuroimaging Initiative (2012) Toward a dynamic biomarker model in Alzheimer's disease. *J Alzheimers Dis* **30**, 91-100.

[35] Samtani MN, Farnum M, Lobanov V, Yang E, Raghavan N, DiBernardo A, Narayan V, the Alzheimer's Disease Neuroimaging Initiative (2012) An improved model for disease progression in patients from the Alzheimer's Disease Neuroimaging Initiative. *J Clin Pharmacol* **52**, 629-644.

[36] Escudero J, Ifeachor E, Zajicek JP, the Alzheimer's Disease Neuroimaging Initiative (2012) Bioprofile analysis: a new approach for the analysis of biomedical data in Alzheimer's disease. *J Alzheimers Dis* **32**, 997-1010.

[37] Jedynak BM, Lang A, Liu B, Katz E, Zhang Y, Wyman BT, Raunig D, Jedynak CP, Caffo B, Prince JL, the Alzheimer's Disease Neuroimaging Initiative (2012) A computational neurodegenerative disease progression score: method and results with the Alzheimer's Disease Neuroimaging Initiative cohort. *Neuroimage* **63**, 1478-1486.

[38] Yang E, Farnum M, Lobanov V, Schultz T, Verbeeck R, Raghavan N, Samtani MN, Novak G, Narayan V, DiBernardo A, the Alzheimer's Disease Neuroimaging Initiative (2011) Quantifying the pathophysiological timeline of Alzheimer's disease. *J Alzheimers Dis* **26**, 745-753.

[39] Cui Y, Sachdev PS, Lipnicki DM, Jin JS, Luo S, Zhu W, Kochan NA, Reppermund S, Liu T, Trollor JN, Brodaty H, Wen W (2012) Predicting the development of mild cognitive impairment: a new use of pattern recognition. *Neuroimage* **60**, 894-901.

[40] Zhang D, Wang Y, Zhou L, Yuan H, Shen D, the Alzheimer's Disease Neuroimaging Initiative (2011) Multimodal classification of Alzheimer's disease and mild cognitive impairment. *Neuroimage* **55**, 856-867.

[41] Koikkalainen J, Pölönen H, Mattila J, van Gils M, Soininen H, Lötjönen J, the Alzheimer's Disease Neuroimaging Initiative (2012) Improved classification of Alzheimer's disease data via removal of nuisance variability. *PLoS One* **7**, e31112.

Supplementary Material

Materials and Methods

Selection of variables

Supplementary Table 1 shows the variables included in the analysis. Details of the data collection protocols in the ADNI are presented in [1].

Supplementary Table 1. Variables used in the analysis

Data modality	Number of items
Mini-Mental State Examination (MMSE)	
All items, except total score	30
Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS)	
All items, except total score	13
Neuropsychological Battery (NeuroBat)	
Clock Drawing Test (all items)	5
Clock Copying Test (all items)	5
Logical Memory Test I and II (all items)	3
Auditory Verbal Learning Test (trials I-V, and list B)	12
Digit Span Test (all items)	4
Category Fluency Test (all items)	6
Trail Making Test (all items)	6
Digit Symbol Substitution Test (all items)	1
Boston Naming Test (all items)	6
Auditory Verbal Learning Test Delayed 30 Minutes (all items)	4
American National Adult Reading Test (all items)	1
Brain volumes from magnetic resonance imaging (MRI)	
Summary measure of total brain parenchyma	1
Total volume of ventricles	1
Volumes of left and right inferior lateral ventricles	2
Volumes of left and right hippocampi	2
Volumes of left and right middle temporal lobes	2
Volumes of left and right inferior temporal lobes	2
Volumes of left and right fusiform gyri	2
Volumes of left and right entorhinal cortices	2
Cerebrospinal fluid samples (CSF)	
Concentration of amyloid-β	1
Concentration of total tau	1
Apolipoprotein E (APOE)	
Type of allele 1 (either ε2, ε3, or ε4)	1

Type of allele 2 (either ε2, ε3, or ε4)	1
Total number of items	114

Examination of the distribution of the regression slopes

Distributions of the slopes from the linear regression of longitudinal Disease State Index values were studied using histograms. On the basis of the histograms, it appeared that the slopes of the stable mild cognitive impairment group may have a bimodal distribution. The expectation maximization (EM) algorithm was used to fit only one Gaussian and a mixture of two Gaussians to the histogram of the slopes. The likelihood ratio test (LRT) was used to test whether the unimodal or bimodal distribution fits better to the data. LRT is defined as

$$LRT = -2ln\left(\frac{L0}{L1}\right) = -2(ln(L0) - ln(L1)), \quad (1)$$

where *ln* is natural logarithm, *L0* is likelihood of the unimodal model, and *L1* is likelihood of the bimodal model. The EM algorithm produced the values of *ln(L0)* and *ln(L1)*. According to simulations done by McLachlan [2], LRT follows the χ^2 distribution with six degrees of freedom when variances of the two normal components in the mixture model are unequal. Thus in this study, six degrees of freedom were used when deciding p-values. If the p-value is below 0.05, unimodal model is rejected and the bimodal model is selected.

Akaike and Bayesian information criteria (AIC and BIC, respectively) were also used for studying the fits of the unimodal and the bimodal distributions. AIC and BIC are methods for comparing fits of a set of models to data. Both methods penalize for increasing the number of estimated parameters, BIC applying a larger penalty term than AIC. The model with the lowest AIC or BIC value is preferred.

Results

The fits of unimodal and bimodal distributions were compared and the results and estimated parameters are shown in Supplementary Table 2. On the basis of AIC, BIC, and LRT, the bimodal distribution fitted better to the slopes of SMCIs than the unimodal distribution.

Supplementary Table 2. Parameter estimates of the unimodal and bimodal Gaussian fit of the regression slopes of the stable mild cognitive impairment group

	Parameter estimates	LL	AIC	BIC	LRT	p
Unimodal normal model		600.6	-1197.3	-1191.3		
(mean ± SD)	0.003 ± 0.004					
Bimodal normal model		628.8	-1247.7	-1232.7	56.4	<0.0005
First mode (mean ± SD)	0.000 ± 0.001					
Second mode (mean ± SD)	0.005 ± 0.004					
Percent in the first mode	59.0 %					

LL, log-likelihood; AIC, Akaike information criterion; BIC, Bayesian information criterion; LRT, likelihood ratio test as defined by Eq. 1; p, p-value determined on the basis of LRT and six degrees of freedom

References

[1] ADNI1 Procedures Manual (2011), http://www.adni-info.org/Scientists/Pdfs/ADNI1_Procedures_Manual_Revised_12052011.pdf, Accessed on January 23, 2012.

[2] McLachlan GJ (1987) On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Appl Statist* **3006**, 318-324.

PUBLICATION III

Multivariate prediction of hippocampal atrophy in Alzheimer's disease

Liedes H., Lötjönen J., Kortelainen J. M., Novak G., van Gils M., Gordon M. F., for the Alzheimer's Disease Neuroimaging Initiative & the Australian Imaging Biomarkers and Lifestyle flagship study of ageing

Journal of Alzheimer's Disease, 2019, 68(4), 1453-1468
https://doi.org/10.3233/JAD-180484

Journal of Alzheimer's Disease 68 (2019) 1453–1468
DOI 10.3233/JAD-180484
IOS Press

Multivariate Prediction of Hippocampal Atrophy in Alzheimer's Disease

Hilkka Liedes[a,*], Jyrki Lötjönen[a,b], Juha M. Kortelainen[a], Gerald Novak[c], Mark van Gils[a], Mark Forrest Gordon[d,e], for the Alzheimer's Disease Neuroimaging Initiative[1] and the Australian Imaging Biomarkers and Lifestyle Flagship Study of Ageing[2]
[a]*VTT Technical Research Centre of Finland Ltd, Tampere, Finland*
[b]*Combinostics Ltd, Tampere, Finland*
[c]*Janssen Pharmaceutical Research and Development, Titusville, NJ, USA*
[d]*Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, CT, USA*
[e]*Current affiliation: Teva Pharmaceuticals, Inc., Frazer, PA, USA*

Accepted 6 February 2019

Abstract.
Background: Hippocampal atrophy (HA) is one of the biomarkers for Alzheimer's disease (AD).
Objective: To identify the best biomarkers and develop models for prediction of HA over 24 months using baseline data.
Methods: The study included healthy elderly controls, subjects with mild cognitive impairment, and subjects with AD, obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI 1) and the Australian Imaging Biomarkers and Lifestyle Flagship Study of Ageing (AIBL) databases. Predictor variables included cognitive and neuropsychological tests, amyloid-β, tau, and p-tau from cerebrospinal fluid samples, apolipoprotein E, and features extracted from magnetic resonance images (MRI). Least-mean-squares regression with elastic net regularization and least absolute deviation regression models were tested using cross-validation in ADNI 1. The generalizability of the models including only MRI features was evaluated by training the models with ADNI 1 and testing them with AIBL. The models including the full set of variables were not evaluated with AIBL because not all needed variables were available in it.
Results: The models including the full set of variables performed better than the models including only MRI features (root-mean-square error (RMSE) 1.76–1.82 versus 1.93–2.08). The MRI-only models performed well when applied to the independent validation cohort (RMSE 1.66–1.71). In the prediction of dichotomized HA (fast versus slow), the models achieved a reasonable prediction accuracy (0.79–0.87).
Conclusions: These models can potentially help identifying subjects predicted to have a faster HA rate. This can help in selection of suitable patients into clinical trials testing disease-modifying drugs for AD.

Keywords: Alzheimer's disease, atrophy, decision support techniques, disease progression, hippocampus, magnetic resonance imaging, regression analysis, statistical models

*Correspondence to: Hilkka Liedes, VTT Technical Research Centre of Finland Ltd, P.O. Box 1300, FIN-33101 Tampere, Finland. Tel.: +358 40 152 6627; Fax: +358 20 722 3499; E-mail: hilkka.liedes@vtt.fi.

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

[2]Data used in the preparation of this article was obtained from the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) funded by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) which was made available at the ADNI database (http://www.loni.usc.edu/ADNI). The AIBL researchers contributed data but did not participate in analysis or writing of this report. AIBL researchers are listed at http://www.aibl.csiro.au.

ISSN 1387-2877/19/$35.00

INTRODUCTION

Alzheimer's disease (AD) is the most common cause of dementia in elderly people and its prevalence is increasing. It has been estimated that 5.4 million Americans have AD in 2016 and the number would increase to 8.4 million by 2030 [1]. Brain pathologies related to AD start to develop already years before the first symptoms appear. AD is characterized by accumulation of extracellular amyloid plaques and intracellular neurofibrillary tangles, neuronal and synaptic loss, and finally atrophy of the brain [2]. The hippocampus is one of the brain regions affected in AD [3]. On the basis of a meta-analysis, the annualized rate of hippocampal atrophy is 1.4% in normal aging, but in AD atrophy is considerably faster with the annualized rate of 4.7% [4].

Prediction of progression in AD is of great interest for clinicians, patients, and researchers. One measure of progression is the conversion from one stage of disease to another, e.g., from mild cognitive impairment (MCI) to dementia. In several studies, hippocampal volume or grading of hippocampal atrophy has been used as a predictor for MCI to AD conversion. These models included, e.g., logistic regression [5–8], support vector machines [7, 9, 10], Naïve Bayes classifier [7], random forest classifier [11, 12], linear discriminant classifier [13], and Disease State Index [7]. In addition, hippocampal volume has been used in prediction of time to conversion using Cox proportional hazards model [5]. These models predicted discrete disease stages, i.e., whether a subject developed AD or not. However, progression of AD is a continuous process and predicting only the conversion is a simplification of the problem.

An alternative method for prediction of discrete disease stages is to predict future scores or change over time in relevant biomarkers. Changes (or future values) in Mini-Mental State Examination (MMSE) and Alzheimer's Disease Assessment Scale – cognitive subscale (ADAS-cog) have been predicted using only baseline data [14–18] or combination of baseline and longitudinal data [10, 19]. Change (or future score) of Clinical Dementia Rating – Sum of Boxes and Global Score have been predicted using several different biomarkers [16, 19, 20].

Performance of cognitive tests may be influenced by a number of factors, including changes in mood, alertness, and fatigue. Moreover, the greatest rate of change in cognitive tests occurs relatively late in the evolution of AD [21]. Because changes in brain structures appear before memory deficits [21], we studied the progression of disease by predicting change in hippocampal volume over time. We selected to focus on hippocampal volume because 1) it is the best established structural biomarker of AD, especially in the early diagnosis [22, 23]; 2) it has already been utilized in clinical trials in AD; 3) it has been qualified for patient enrichment in pre-dementia trials by the European Medicines Agency [24] and the Food and Drug Administration has issued a Letter of Support for the same purpose [25]. By developing prediction models for change in hippocampal volume, we might be able to identify subjects at the time of a single baseline study as at risk of more rapid atrophy rate and faster progression of the disease. This information could be useful, e.g., in selection of suitable patients for clinical trials of potential disease-modifying drugs.

The objective of this study was to develop a multivariate model for predicting hippocampal atrophy rate at month 24 using data measured at the baseline. First, models with a full set of variables including cognitive and neuropsychological tests, cerebrospinal fluid samples (CSF), genetic information, and features extracted from magnetic resonance imaging (MRI) were developed and evaluated. Second, the performance of models including only MRI features was compared to the models including the full set of variables using the same subjects who had all variables available. Development of the models with only MRI data is beneficial because not all variables are always available, and different neuropsychological tests can be used in different centers. Third, performance of the models including only MRI features was evaluated with a larger population having MRI, regardless of availability of other variables. Finally, the models including only MRI features were evaluated using an independent validation data set. The models with the full set of variables could not be evaluated with the independent cohort due to the limited availability of the needed data modalities in the validation data set.

MATERIALS AND METHODS

Study population

ADNI 1

Data used in the analyses were obtained from the Alzheimer's Disease Neuroimaging Initiative, ADNI 1 database (http://adni.loni.usc.edu). The ADNI 1 was launched in 2003 as a public-private partnership, led by Principal Investigator Michael Weiner. ADNI 1 aimed at testing whether a combination of imaging,

biological markers, and clinical and neuropsychological assessments can measure the progression of MCI and AD. ADNI 1 has been followed by ADNI GO, ADNI 2, and ADNI 3 studies, in which new subjects were recruited and follow-up of a group of the already included subjects was continued. Up-to-date information about all ADNI studies can be found at http://www.adni-info.org.

Approximately 800 subjects of age 50 to 90 years have been recruited at around 50 sites in the United States and Canada into ADNI 1. The goal of the ADNI 1 was to recruit 200 healthy elderly controls (NC), 400 subjects with MCI, and 200 subjects with early AD. The subjects underwent cognitive assessment, neuropsychological testing, and 1.5 T MRI at intervals of six or twelve months for two to four years. Other tests, such as fluorodeoxyglucose positron emission tomography (FDG-PET), 3 T MRI, and blood and CSF sampling, were performed less frequently or only in a portion of the subjects [26]. The data were downloaded from the ADNI web site on 22 August in 2014.

AIBL

The purpose of the Australian Imaging Biomarkers and Lifestyle Flagship Study of Ageing (AIBL) was to discover biomarkers, cognitive characteristics, and health and lifestyle factors that affect development of symptomatic AD. The study was launched in 2006 and it has collected longitudinal data for over 4.5 years. Its aim was to recruit at least 1000 participants of at least 60 years. The subjects consisted of healthy volunteers, subjects with MCI, and subjects with AD. Data were collected in Perth in Western Australia and Melbourne in Victoria (http://aibl.csiro.au). Data were collected by the AIBL study group. AIBL study methodology has been reported previously [27].

AIBL has a collaborative agreement with the US-based Alzheimer's Association (http://www.aibl.csiro.au/adni/index.html). Thus, data of the AIBL subjects having PET and MRI are available through the same technical infrastructure as ADNI. AIBL subjects with MRI and PET data constitute about 25% of the full AIBL cohort. Some of the oldest baseline AIBL data were acquired without ADNI-compliant MRI sequences and these data have not been provided to the ADNI portal (http://www.aibl.csiro.au/adni/imaging.html). AIBL data were downloaded from the ADNI web site on 27 November 2014.

Predictor variables and included subjects

In order to develop and validate the models for prediction of hippocampal atrophy rate, four sets of analyses were conducted, and they are summarized in Table 1. **Analysis 1** included the full set of variables: cognitive and neuropsychological tests (MMSE, ADAS-cog, Clinical Dementia Rating (CDR), clock draw and clock copy, Auditory Verbal Learning Test (AVLT), Digit Span, Category Fluency Test, Trail Making Test, Digit Symbol Substitution Test, Boston Naming Test, Auditory Verbal Learning Test Delayed 30 Minutes, American National Adult Reading Test), apolipoprotein E (APOE), CSF (tau, *p*-tau, amyloid-β), and features extracted from MRI (explained in the section *Cross-sectional image quantification*). Only a subset of subjects in ADNI 1 had CSF biomarkers available, thus, this analysis was restricted to the subjects having CSF biomarkers at the baseline and MRI at the baseline and month 24 (N = 281). Availability of other variables did not affect the inclusion of a subject into the analysis, meaning that the subject was included even if she/he had missing values in other variables. Purpose of the **Analysis 2** was to compare the models including only MRI features to the models including the full set of variables. Thus, it included only MRI features as predictors and contained ADNI 1 subjects with baseline and 24-month MRI, regardless of availability of CSF biomarkers (N = 530). In order to get comparable results with Analysis 1, this population was divided into cross-validation folds and the results

Table 1
Summary of the analyses

	Analysis 1	Analysis 2	Analysis 3	Analysis 4
Data	ADNI 1	ADNI 1	ADNI 1	ADNI 1, AIBL
Predictors	Cognitive tests, neuropsychological tests, CSF, APOE, MRI	MRI	MRI	MRI
Validation	10-fold cross-validation	10-fold cross-validation using 530 subjects	10-fold cross-validation	Train: ADNI 1 Test: AIBL
Number of subjects	281	Results are shown for 281 subjects	530	ADNI 1 = 530 AIBL = 176

were calculated using the same subjects as in Analysis 1 (N = 281). **Analysis 3** was otherwise similar to Analysis 2, but the results were calculated using all 530 subjects. **Analysis 4** evaluated generalizability of the MRI-only models with an independent test data. The models were trained using the same ADNI 1 population as in Analysis 3 and tested using the AIBL population having the baseline and 18-month MRI available (N = 176). Evaluation with AIBL was done only for the MRI-only models because the full set of all variables was not available in AIBL.

All predictor variables are listed in Supplementary Table 1. The full set included in total 534 variables, of which 422 were MRI features. Missing values in the predictor variables were imputed with the medians. The effect of imputation on the results should be subtle, as 99.1%, 99.8%, and 100% of the variables had at most three missing values in Analysis 1 and 2, Analysis 3, and Analysis 4, respectively. This study did not include amyloid PET or FDG-PET because they were performed only in a subset of patients and a requirement to have all data modalities (baseline and follow-up MRI, CSF, and PET) available would have substantially reduced the number of eligible subjects.

In this study, subjects were considered as NC, stable MCI (SMCI), or AD if their diagnosis stayed as NC, MCI, or AD, respectively, for 36 months. If the diagnosis changed from MCI to AD during 36 months, a subject was considered as progressive MCI (PMCI). If diagnoses changed otherwise than from MCI to AD (e.g., NC $\longrightarrow$ AD or NC $\longrightarrow$ MCI $\longrightarrow$ NC), the diagnosis was labeled as unknown. The subjects with unknown diagnoses were also included into the analyses. A limit of 36 months was selected because it was the latest time point on which diagnoses were available in both ADNI 1 and AIBL.

Table 2 shows demographics of the study populations in the different analyses. No statistically significant differences were observed in any demographic variables between 281 subjects in Analysis 1 and 2 and 530 subjects in Analysis 3. The subjects in ADNI 1 were older, had lower MMSE total score and lower hippocampal volume at the baseline than the subjects in AIBL. ADNI 1 included more males than females whereas AIBL had equal amount of both genders. Proportions of the subjects in the different diagnostic classes were not equal between ADNI 1 and AIBL: the number of subjects in each diagnostic class was roughly equal in ADNI 1 while the majority of the subjects were healthy controls in AIBL (approximately 10% were subjects with MCI, and 12% were subjects with AD). Amyloid positivity was defined on the basis of amyloid-β levels, using a threshold of 192 pg/ml [28]. In Analysis 1 and 2, 69% of the subjects were amyloid positive. Due to lack of CSF samples, amyloid positivity is not reported for Analysis 3 and 4. However, the proportion of amyloid

Table 2
Demographics of the study population at the baseline

	ADNI 1 Analysis 1&2	Analysis 3	AIBL Analysis 4	*p* Analysis 1&2 versus 3	*p* Analysis 1&2 versus 4	*p* Analysis 3 versus 4
N	281	530	176			
Age (y)	75.1 ± 6.7	75.4 ± 6.5	71.9 ± 7.2	0.633	<0.0001	<0.0001
Gender				0.756	0.048	0.054
Male	167 (59)	309 (58)	88 (50)			
Female	114 (41)	221 (42)	88 (50)			
MMSE	26.8 ± 2.6	27.0 ± 2.6	27.8 ± 2.9	0.268	<0.0001	<0.0001
HC (mm^3)	3826 ± 659	3837 ± 676	4185 ± 582	0.985	<0.0001	<0.0001
Diagnosis				0.968	<0.0001	<0.0001
NC	75 (27)	151 (28)	119 (68)			
SMCI	61 (22)	118 (22)	12 (7)			
PMCI	64 (23)	118 (22)	6 (3)			
AD	63 (22)	109 (21)	21 (12)			
Unknown	18 (6)	34 (6)	18 (10)			
Aβ	169 ± 56	NA	NA			
Aβ+	193 (69)	NA	NA			

N, number of subjects; MMSE, total score of Mini-Mental State Examination; HC, hippocampal volume (left + right); NC, healthy elderly control; SMCI, stable mild cognitive impairment; PMCI, progressive mild cognitive impairment; AD, Alzheimer's disease; Unknown, diagnosis changed otherwise than from MCI to AD; Aβ, amyloid-β; Aβ+, number of amyloid positive subjects, i.e., subjects whose amyloid-β was below 192 pg/ml [28]; NA, not available due to lack of cerebrospinal fluid samples. Values are expressed as mean ± standard deviation or as count (percentage).

positive subjects in AIBL has been reported elsewhere: Villemagne et al. [29] reported that 26% of NC, 67% of MCI and 95% of AD cases were amyloid positives; corresponding numbers from Rowe et al. [30] were 33%, 68%, and 98%. These numbers were based on standardized uptake value ratio of ^{11}C-Pittsburgh Compound B in PET imaging. Demographics stratified by the diagnostic classes are presented in Supplementary Table 2.

Cross-sectional image quantification

This section describes calculation of MRI features used as predictors in the modelling. Volumetry, tensor-based morphometry (TBM), and voxel-based morphometry (VBM) were used to extract the features from the baseline MRI [31]. Even though these features correlate with each other, all of them were included because they provide complementary information on brain health. In addition, the selected prediction methods perform feature selection which will reduce the number of features (see the section *Modelling methods*).

Volumetry

The volumetric features were calculated using the multi-atlas segmentation method extended by the expectation maximization classification [32]. Segmentation was performed using the Neuromorphometrics atlases (http://www.neuromorphometrics.com/), including altogether 139 brain regions. Figure 1 shows an example of the segmentation. Volume of hippocampus was also computed using another atlas from ADNI [32].

Tensor-based morphometry

In TBM, a reference image is registered to the patient image. When the reference image is registered similarly to images from different individuals, the local volume changes can be compared across different groups of individuals, e.g., patients with AD and healthy controls. For example, the size of the lateral ventricles tends to increase in AD which can be observed by the local volume increase in that area when compared with healthy controls. In this study, TBM features were computed by integrating the local volume changes over the regions of interests obtained in volumetric segmentation (see the section *Volumetry* and Supplementary Table 1). The method presented by Koikkalainen et al. [33] was used.

Voxel-based morphometry

While TBM measures the local volume changes, VBM quantifies the local concentration of the gray matter (GM) after accounting for global differences in anatomy by registering a patient image to a reference image [34]. As in TBM, the process was performed on images from a high number of individuals and the VBM features were obtained by integrating the local concentrations over the regions of interest from the volumetric segmentation (see the section *Volumetry* and Supplementary Table 1).

Longitudinal image quantification

This section describes calculation of the observed change in total hippocampal volume (average atrophy over the left and right) which was used as a reference value in the prediction. Change in

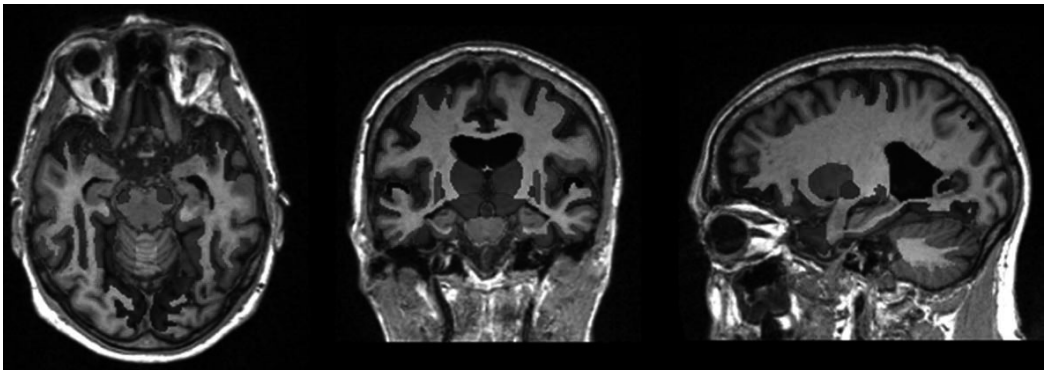


Fig. 1. Example of the brain structure segmentation where each segmented brain structure region is shown in different colors. (Color figure available in the online version).

the hippocampal volume was estimated using the extended boundary shift integral (eBSI) method [35]. Longitudinal eBSI was used instead of the cross-sectional volumetric method described in the section *Volumetry*. The longitudinal method utilizes several images from different time points simultaneously in the calculation of atrophy rate, whereas images from different time points are processed separately in the cross-sectional analysis. Longitudinal methods provide typically more accurate measures of atrophy rate than cross-sectional methods.

The original BSI method is a semi-automatic approach which estimates the loss of GM either locally or globally in the border of CSF and GM [36]. The atrophy rate is computed by integrating the intensity differences between baseline and follow-up images in a certain boundary region. The extended version of the BSI used in this study integrates over tissue or structure probabilities instead of integrating over intensity changes as is done in the original BSI [35].

ADNI 1 included follow-up MRI at months 12, 24, and at the later time points and AIBL at month 18. The annual atrophy rate was calculated from 24-month images using ADNI 1 and from 18-month images using AIBL. Atrophy rate in AIBL was multiplied by the factor of 2/3 to make it correspond with the annual atrophy rate in ADNI 1.

Modelling methods

Two models were used in this study to predict hippocampal atrophy rate on the basis of the data measured at the baseline: 1) regularized least mean square regression with the elastic net regularization and 2) least absolute deviation regression with regularized least mean square regression as a feature selection method.

Regularized least mean square regression

The normal linear least mean square regression (LMS) includes all variables in the model, which is not very practical when the number of predictor variables is high. An alternative for normal LMS regression is the regularized least mean square regression with the elastic net regularization defined as

$$\min_{\beta_0,\,\beta}\left(\frac{1}{2N}\sum_{i=1}^{N}\left(y_i-\beta_0-x_i^T\beta\right)^2+\lambda P_\alpha(\beta)\right), \quad (1)$$

where

$$P_\alpha(\beta)=\frac{(1-\alpha)}{2}\|\beta\|_2^2+\alpha\|\beta\|_1=\sum_{j=1}^{p}\frac{(1-\alpha)}{2}\beta_j^2+\alpha\left|\beta_j\right|, \quad (2)$$

where β_0 is intercept, β is vector of regression coefficients, N is number of subjects, y_i is the response (observed outcome) of the subject i, x_i is the vector including predictor data for the subject i, λ is a positive regularization parameter, α has values between 0 and 1 defining the weight of L1 and L2 norms, p is the number of predictors in the model [37, 38].

Regularized LMS regression with elastic net regularization includes a penalty term $P(\alpha)$ that constrains the size of the estimated regression coefficients, thus, it sets some of the regression coefficients to zero leading in practice to feature selection. The penalty term includes both L1 and L2 norms of the regression coefficients. When the parameter α has a value of one, the L1 norm is used and elastic net is the same as lasso regularized regression. When the parameter α is close to zero, the L2 norm is emphasized and the model approaches ridge regression. For other values of α, the penalty term is a weighted combination of L1 and squared L2 norms. Regularization parameter λ defines the amount of regularization. When λ increases, the number of non-zero regression coefficients decreases. An optimal value for λ was defined using the nested cross-validation, i.e., the training population was further divided into five folds and cross-validation was performed over these folds. The largest λ providing the mean squared error (MSE) within one standard error of the minimum MSE was selected. Once the λ was selected, the whole training population of the fold was used to define optimal regression coefficients.

Least absolute deviation regression

Least absolute deviation regression (LAD) is an alternative method for the least mean square regression [39]. It minimizes the absolute values of the residuals defined as

$$\min_{\beta_0,\,\beta}\left(\sum_{i=1}^{N}\left|y_i-\beta_0-x_i^T\beta\right|\right), \quad (3)$$

where β_0 is intercept, β is vector of regression coefficients, N is number of subjects, y_i is the response (observed outcome) of the subject i, x_i is the vector including predictor data for the subject i.

LAD is more robust against outliers in the data than the LMS regression. Feature selection is not embedded in the LAD regression. In this work, features having non-zero coefficients in the regularized LMS regression with the elastic net regularization were selected into the LAD.

Effect of nuisance variables

The nuisance variables affect the analysis results, but they are not of primary interest as classification features. Demographics like age, gender, education, and weight have been shown to interact with neuropsychological assessments and biomarkers [40–42]. Age and gender were available both in ADNI 1 and AIBL studies, thus, their effect on predictors was removed using the method described by Koikkalainen et al. [43]. The method is based on the data variability in the population of healthy controls because data from this group should not contain any disease-related variability. First, linear absolute deviation regression between the predictor and the nuisance variables was performed in healthy controls of the training population. Then, corrected values for other subjects were calculated using the regression coefficients:

$$c_{corr} = c_{orig} - \left(\beta_0 + \beta_{age} \times age + \beta_{gender} \times gender\right), \quad (4)$$

where c_{corr} is the corrected value of the predictor, c_{orig} is the original value of the predictor, b_0 is the intercept of the regression between the predictor and nuisance features, and β_{age}, β_{gender}, are regression coefficients for age and gender, respectively.

Performance evaluation

In Analysis 1, 2, and 3, prediction performance of the regularized LMS regression and LAD regression models was evaluated using the 10-fold cross validation stratified according to the diagnostic classes of NC, SMCI, PMCI, and AD. Stratified means that the proportions of the diagnostic classes were roughly equal in each cross-validation fold. In Analysis 4, generalizability of the models was studied with independent validation data by training the models with the whole ADNI 1 cohort and testing them with the AIBL cohort. Root-mean-square errors (RMSE) and Spearman correlation coefficients were used as performance measures. In addition, it was studied how well the models predict which of the subjects have fast or slow rate of hippocampal atrophy, i.e., subjects whose atrophy rate is less or more than a certain limit. One option to define the limit is to use middle point between the average atrophy rate in NC and AD. Instead of using the atrophy rates for NC and AD from Barnes et al. [4], averages were calculated in cross-validation using our study population. Accuracy, sensitivity, specificity, positive predictive value, and negative predictive value were reported from this analysis. Performance and validity of the models was visually inspected by plotting observed versus predicted hippocampal atrophy rates. Residuals of the models were also studied. Wilcoxon singed-rank test was used to study whether the differences in RMSEs and Spearman correlation coefficients were statistically significant between the full models (Analysis 1) and MRI-only models (Analysis 2).

In ADNI 1, the number of cases in each diagnostic group was roughly equal while AIBL contained nearly six times more healthy controls than subjects with AD (Table 2). This has implications especially for calculation of the correlation coefficient. The imbalance was taken into account by choosing randomly 20 healthy controls with all subjects in other diagnostic groups and then calculating the performance measures. This number was selected because it was close to the number of AD cases in the AIBL (N = 21). This process was repeated 20 times.

Other statistical tests

Demographics of the study populations in the different analyses were compared using Mann-Whitney *U* test (continuous variables) or Chi-squared test (categorical variables). Hippocampal atrophy rates in the different diagnostic classes was compared using Kruskal-Wallis test with Dunn's *post hoc* test. In all analyses, significance threshold was set at 0.05. The analyses were performed in MATLAB R2015b (The MathWorks, Inc., Natick, MA, USA) or in IBM SPSS Statistics version 22 (IBM, Armonk, NY, USA).

RESULTS

Distribution of changes in hippocampal volumes

Figure 2 and Table 3 show distributions of annual hippocampal atrophy rates in the different diagnostic classes. In all analyses of ADNI 1, NCs had smaller annual atrophy rates than subjects in other classes ($p < 0.0005$), also SMCIs had smaller annual atrophy rates than PMCIs or ADs ($p < 0.0005$). There were no statistically significant differences between PMCIs and ADs ($p = 0.329$ in Analysis 1 & 2, $p = 0.068$ in Analysis 3). Also in AIBL, NCs had smaller

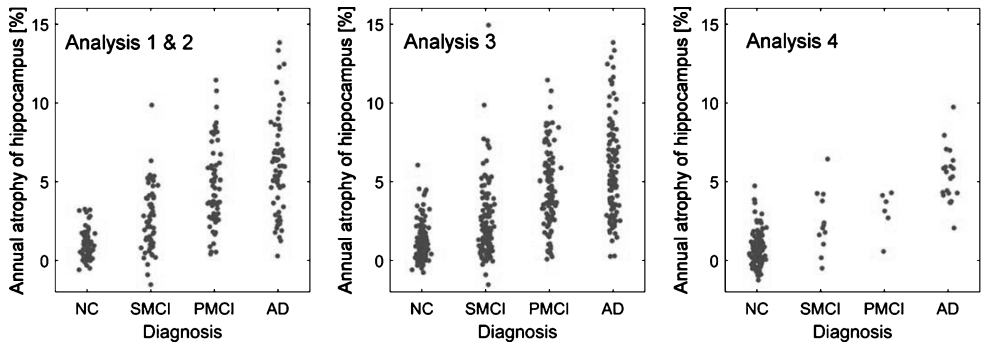


Fig. 2. Distribution of annual decrease in the hippocampal volume in the different subject groups. Analysis 1 & 2 included ADNI 1 subjects with CSF available. Analysis 3 included ADNI 1 subjects with baseline and follow-up MRI, regardless of availability of CSF. Analysis 4 included AIBL subjects. NC, normal control; SMCI, stable mild cognitive impairment; PMCI, progressive mild cognitive impairment; AD, Alzheimer's disease.

Table 3
Summary of annual decrease (unit [%]) of the hippocampal volume in different subject groups

	ADNI 1		AIBL
	Analysis 1 & 2	Analysis 3	Analysis 4
NC	$1.06 \pm 0.87^{b,c,d}$	$1.17 \pm 1.11^{b,c,d}$	$0.77 \pm 1.09^{b,c,d}$
SMCI	$2.58 \pm 2.02^{a,c,d}$	$2.47 \pm 2.30^{a,c,d}$	2.47 ± 1.93^{a}
PMCI	$4.63 \pm 2.55^{a,b}$	$4.43 \pm 2.36^{a,b}$	3.10 ± 1.38^{a}
AD	$6.06 \pm 3.08^{a,b}$	$5.84 \pm 2.97^{a,b}$	5.35 ± 1.70^{a}

Analysis 1 included cognitive and neuropsychological tests, CSF, APOE, and MRI; Analysis 2, 3, and 4 included only MRI. NC, normal control; SMCI, stable mild cognitive impairment; PMCI, progressive mild cognitive impairment; AD, Alzheimer's disease; ^{a,b,c,d}significantly ($p < 0.05$) different from NC, SMCI, PMCI, AD, respectively. Values are expressed as mean ± standard deviation.

annual atrophy rate than subjects in other three classes ($p = 0.017$ in comparison to SMCI, $p = 0.020$ to PMCI, $p < 0.0005$ to AD). There were no statistically significant differences between other classes. The mean annual atrophy rates follow the numbers reported in a meta-analysis for normal aging and Alzheimer's disease: 1.4 % and 4.7 %, respectively [4]. Absolute loss of hippocampal volume is shown in Supplementary Table 3.

Prediction performance

Table 4 shows RMSEs and Spearman correlation coefficients of the models. Figure 3 presents scatter plots of the observed and predicted hippocampal atrophy rates for the models with both L1 and L2 norms. Corresponding plots with only L1 norm are presented in Supplementary Figure 1. As expected, the models including cognitive and neuropsychological tests, CSF, APOE, and MRI as input variables performed better than the models including only MRI features: RMSEs were smaller and correlation coefficients were higher (all $p < 0.05$). The predicted atrophy rates also corresponded better to the observed atrophy rates in the full model than in the MRI-only model. Especially, at the higher atrophy rate levels, the MRI-only models seemed to underestimate the real change to a greater degree than the full model, which also showed some underestimation (Fig. 3).

Table 5 presents the performance of the models to predict which of the subjects have fast or slow rate of hippocampal atrophy. The limit for the fast atrophy rate was set to average of atrophy rates in NC and AD using our study population. In all analyses, the accuracy was reasonable (0.79–0.87). The models with the full set of variables provided similar accuracy, sensitivity, and specificity values. The MRI-only models had lower sensitivities than specificities, which is in line with the result that the MRI-only models underestimated the real atrophy rate.

In an attempt to enhance the MRI-only models, quadratic terms were added to these models; however, the changes in the performance were subtle (results not shown here). Furthermore, addition of quadratic terms did not reduce the problem of underestimation at the higher atrophy rates. Thus, it was decided

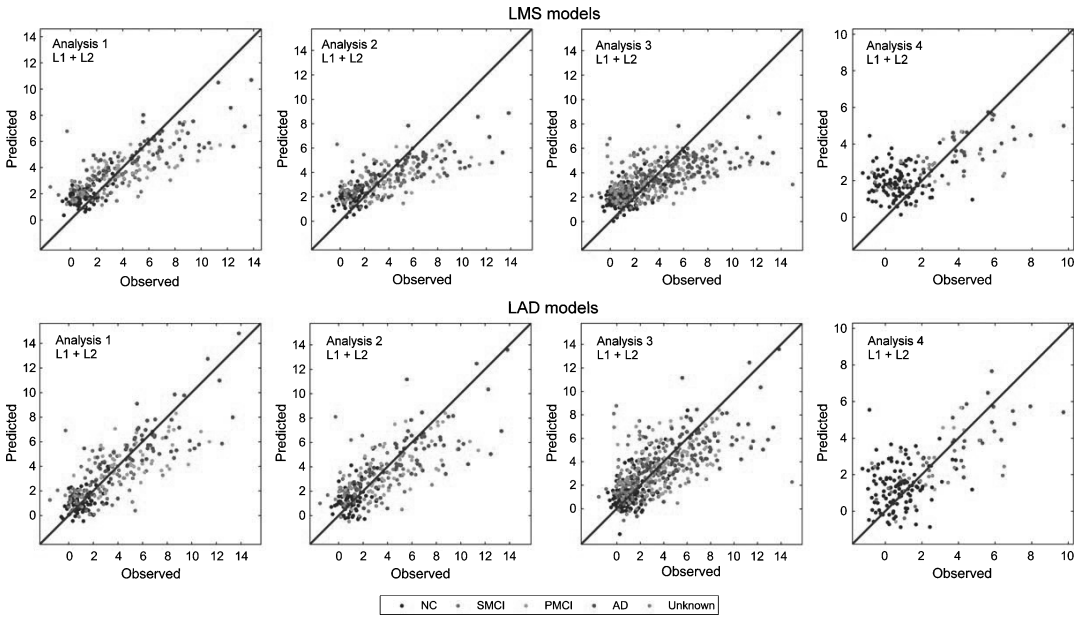


Fig. 3. Observed and predicted annual hippocampal atrophy rates for the LMS (top row) and LAD (bottom row) models with the L1 + L2 norm in the regularization. LMS, least mean square regression; LAD, least absolute deviation regression; NC, normal control; SMCI, stable mild cognitive impairment; PMCI, progressive mild cognitive impairment; AD, Alzheimer's disease; Unknown, diagnosis of the subject changed otherwise than from MCI to AD. (Color figure available in the online version).

Table 4
Prediction performance of the models

Model	Feature selection	α	#Features		RMSE [%]		Spearman rho		*p*	
Analysis 1: full model + cross-validation with ADNI 1 (N = 281)										
LMS	L1 + L2	0.5	27	(2)	1.79	(0.30)**	0.78	(0.06)**	<0.0001	(0.0001)
LMS	L1	1.0	16	(4)	1.82	(0.30)**	0.77	(0.07)*	<0.0001	(0.0001)
LAD	L1 + L2	0.5	27	(2)	1.76	(0.34)*	0.76	(0.09)*	0.0001	(0.0001)
LAD	L1	1.0	16	(4)	1.76	(0.33)*	0.77	(0.09)**	0.0001	(0.0002)
Analysis 2: MRI-only model + cross-validation with ADNI 1 (N = 281)										
LMS	L1 + L2	0.5	18	(4)	2.06	(0.35)	0.72	(0.09)	0.0005	(0.0014)
LMS	L1	1.0	10	(2)	2.08	(0.36)	0.72	(0.09)	0.0004	(0.0008)
LAD	L1 + L2	0.5	18	(4)	1.93	(0.44)	0.72	(0.08)	0.0003	(0.0006)
LAD	L1	1.0	10	(2)	1.93	(0.40)	0.71	(0.09)	0.0005	(0.0012)
Analysis 3: MRI-only model + cross-validation with ADNI 1 (N = 530)										
LMS	L1 + L2	0.5	18	(4)	2.11	(0.33)	0.68	(0.07)	<0.0001	(<0.0001)
LMS	L1	1.0	10	(2)	2.12	(0.33)	0.68	(0.08)	<0.0001	(<0.0001)
LAD	L1 + L2	0.5	18	(4)	2.07	(0.38)	0.68	(0.08)	<0.0001	(<0.0001)
LAD	L1	1.0	10	(2)	2.06	(0.34)	0.67	(0.07)	<0.0001	(<0.0001)
Analysis 4: MRI-only model + validation with AIBL (N = 176)										
LMS	L1 + L2	0.5	22		1.71	(0.06)	0.71	(0.04)	<0.0001	(<0.0001)
LMS	L1	1.0	20		1.71	(0.07)	0.69	(0.04)	<0.0001	(<0.0001)
LAD	L1 + L2	0.5	22		1.66	(0.07)	0.71	(0.03)	<0.0001	(<0.0001)
LAD	L1	1.0	20		1.67	(0.07)	0.71	(0.03)	<0.0001	(<0.0001)

LMS, least mean square regression; LAD, least absolute deviation regression: L1, L1 norm used in the regularization; L1 + L2, L1 and L2 norms used in the regularization; α, parameter for controlling weights of L1 and L2 norms in the regularization; RMSE, root mean square error; *p*, *p*-value of Spearman correlation coefficient (rho). For ADNI 1, values are presented as mean (standard deviation) over cross-validation folds. Imbalance in the number of subjects in the different diagnostic groups in AIBL was accounted by randomly selecting 20 healthy controls together with all subjects from other groups. This was repeated multiple times and values are presented as mean (standard deviation) over these iterations. * and **: *p*-value <0.05 and <0.01, respectively, when compared to the corresponding model in Analysis 2 (Wilcoxon signed-rank test).

Table 5
Performance of the models to predict rate of hippocampal atrophy, dichotomized as fast or slow

Model	Feature selection	α	Accuracy		Sensitivity		Specificity		PPV		NPV	
Analysis 1: full model + cross-validation with ADNI 1 (N = 281)												
LMS	L1 + L2	0.5	0.84	(0.08)	0.85	(0.14)	0.84	(0.07)	0.77	(0.10)	0.89	(0.11)
LMS	L1	1.0	0.83	(0.06)	0.82	(0.14)	0.85	(0.07)	0.78	(0.10)	0.88	(0.10)
LAD	L1 + L2	0.5	0.83	(0.08)	0.84	(0.14)	0.83	(0.07)	0.75	(0.11)	0.88	(0.11)
LAD	L1	1.0	0.83	(0.08)	0.85	(0.15)	0.83	(0.08)	0.76	(0.12)	0.89	(0.11)
Analysis 2: RI-only model + cross-validation with ADNI 1 (N = 281)												
LMS	L1 + L2	0.5	0.82	(0.10)	0.77	(0.13)	0.85	(0.09)	0.78	(0.13)	0.84	(0.12)
LMS	L1	1.0	0.82	(0.06)	0.79	(0.13)	0.84	(0.08)	0.77	(0.10)	0.85	(0.11)
LAD	L1 + L2	0.5	0.82	(0.08)	0.78	(0.10)	0.85	(0.10)	0.79	(0.11)	0.85	(0.08)
LAD	L1	1.0	0.82	(0.10)	0.76	(0.14)	0.86	(0.10)	0.79	(0.13)	0.84	(0.12)
Analysis 3: MRI-only model + cross-validation with ADNI 1 (N = 530)												
LMS	L1 + L2	0.5	0.79	(0.06)	0.73	(0.12)	0.82	(0.06)	0.71	(0.09)	0.83	(0.09)
LMS	L1	1.0	0.79	(0.05)	0.73	(0.11)	0.82	(0.06)	0.71	(0.09)	0.83	(0.08)
LAD	L1 + L2	0.5	0.79	(0.04)	0.73	(0.09)	0.83	(0.08)	0.72	(0.10)	0.84	(0.06)
LAD	L1	1.0	0.79	(0.06)	0.71	(0.12)	0.83	(0.07)	0.72	(0.09)	0.83	(0.08)
Analysis 4: MRI-only model + validation with AIBL (N = 176)												
LMS	L1 + L2	0.5	0.87	(0.01)	0.72	(0.01)	0.97	(0.01)	0.94	(0.02)	0.84	(0.01)
LMS	L1	1.0	0.85	(0.02)	0.69	(0.01)	0.97	(0.02)	0.93	(0.04)	0.82	(0.02)
LAD	L1 + L2	0.5	0.87	(0.01)	0.72	(0.01)	0.97	(0.01)	0.93	(0.02)	0.84	(0.02)
LAD	L1	1.0	0.86	(0.01)	0.76	(0.01)	0.93	(0.01)	0.87	(0.02)	0.85	(0.01)

The limit for fast hippocampal atrophy was based on the mean of NC and AD. The limit was on average 3.6 in Analysis 1 and 3.5 in rest of the analyses. PPV, positive predictive value; NPV, negative predictive value; LMS, least mean square regression; LAD, least absolute deviation regression: L1, L1 norm used in the regularization; L1 + L2, L1 and L2 norms used in the regularization; α, parameter for controlling weights of L1 and L2 norms in the regularization. For ADNI 1, values are presented as mean (standard deviation) over cross-validation folds. Imbalance in the number of subjects in the different diagnostic groups in AIBL was accounted by randomly selecting 20 healthy controls together with all subjects from other groups. This was repeated multiple times and the values are presented as mean (standard deviation) over these iterations for AIBL.

to move forward with the simpler models with only linear terms.

In addition, we re-ran the analyses using only neuropsychological and cognitive tests, CSF, and APOE, excluding the MRI features, to evaluate their effect on the results. These non-MRI models performed worse than the full or MRI-only models. Their RMSEs were 2.18–2.24, Spearman rhos were 0.68–0.71, and accuracies in the prediction of fast atrophy rate were 0.79–0.81 (results not shown in the tables or figures). This is a result that can be reasonably expected, as the atrophy of the brain usually begins before performance in neuropsychological and cognitive testing deteriorates.

The MRI-only models generalized well across the independent validation cohort. RMSEs were smaller in validation with AIBL than in other three analyses (Table 4). Accuracies for predicting fast atrophy rate were also higher in AIBL than in ADNI 1; however, this was mostly due to the high specificity values (Table 5).

The LMS models performed worse than LAD models in all analyses. This can clearly be seen in Fig. 3. The predicted atrophy rates of the LAD models corresponded better to the observed atrophy rates. The LMS models seemed to overpredict change in subjects with low hippocampal atrophy rates and underpredict at high rates. There was only a minor difference between the LAD models including both L1 and L2 norms and the models including only L1 norm.

Selected variables

Linear regression with elastic net regularization works as a feature selection method because it sets some of the regression coefficients to zero. Table 6 presents variables selected into the models in Analysis 1 including the full set of variables. In total, 19 variables were selected in all cross-validation folds when both L1 and L2 norms were used in the regularization. Recalls of words (delayed recall from MMSE, word recall and delayed word recall from ADAS-cog, and trials 3 and 4 from AVLT), orientation and modified total score from ADAS-cog were selected from the neuropsychological and cognitive tests. Amyloid-β, *p*-tau, and presence of APOE ε4 were also selected. Volumes of hippocampus,

Table 6
Variables selected by the feature selection in Analysis 1, showing only the variables selected at least in five folds

L1 + L2 (α = 0.5)	*N*	L1 (α = 1.0)	*N*
MMSE: delayed recall ball	10	ADAS-cog: modified total score	10
ADAS-cog: Q1 word recall	10	APOE: presence of ε4	10
ADAS-cog: Q4 delayed word recall	10	Volume of left hippocampus	10
ADAS-cog: Q7 orientation	10	Volume of right inferior lateral ventricle	10
ADAS: modified total score	10	TBM: left hippocampus	10
AVLT: trial 3 total	10	CSF: *p*-tau	9
AVLT: trial 4 total	10	VBM: left hippocampus	9
CSF: amyloid beta	10	VBM: left inferior lateral ventricle	9
CSF: *p*-tau	10	MMSE: delayed recall ball	8
APOE: presence of ε4	10	Volume of left inferior lateral ventricle	8
Volume of left hippocampus	10	CSF: amyloid beta	7
Volume of right hippocampus	10	ADAS-cog: Q7 orientation	6
Volume of left inferior lateral ventricle	10	Volume of right hippocampus	6
Volume of right inferior lateral ventricle	10	VBM: global	6
TBM: left hippocampus	10	ADAS-cog: Q1 word recall	5
TBM: right inferior lateral ventricle	10	AVLT: trial 4 total	5
VBM: left amygdala	10	TBM: right inferior lateral ventricle	5
VBM: left hippocampus	10		
VBM: left inferior lateral ventricle	10		
MMSE: total score	9		
CDR: sum of boxes	9		
TBM: left amygdala	8		
VBM: global	8		
CDR: home and hobbies	7		
AVLT: trial 5 total	6		
ADAS-cog: Q8 word recognition	5		
CSF: tau	5		

L1 + L2, L1 and L2 norms used in the regularization; L1, L1 norm used in the regularization; α, parameter for controlling weights of L1 and L2 norms; *N*, number of times a variable was selected in different cross-validation folds; MMSE, Mini-Mental State Examination; ADAS-cog, Alzheimer's Disease Assessment Scale – cognitive subscale; CDR, Clinical Dementia Rating; AVLT, Auditory Verbal Learning Test; CSF, cerebrospinal fluid; APOE, apolipoprotein E; VBM, voxel-based morphometry; TBM, tensor-based morphometry.

inferior lateral ventricles, and amygdala calculated using different methods were selected from the pool of MRI features. Other variables were selected less frequently. When the L1 norm was used in the regularization, fewer variables were selected and they included partly the same variables as in L1 + L2 regularization. Modified total score of ADAS-cog, presence of APOE ε4, hippocampus, and inferior lateral ventricles were selected in all folds.

Table 7 shows variables selected into the MRI-only models when the models were trained with the whole ADNI 1 population and tested with AIBL. When L1 and L2 norms were used in the regularization, 22 features were selected. In total 12 features described volumes of hippocampus, inferior lateral ventricles, and amygdala calculated using different methods. The rest of the selected features described other brain areas. The regularization with L1 norm selected 20 features and all of them, except the right parahippocampal gyrus from VBM, were the same as in the regularization with L1 and L2 norms.

DISCUSSION

Atrophy of the hippocampus is one of the biomarkers for AD [3]. In this study, we developed multivariate models for prediction of hippocampal atrophy at month 24 using variables measured at the baseline. The models were based on regularized least mean square regression and least absolute deviation regression. Performance of the models including variables from different data modalities, like neuropsychological and cognitive tests, CSF biomarkers, MRI features, and APOE, was compared to the models including only MRI features. Performance of the MRI-only models was also evaluated using an independent validation data set.

The models comprising different data modalities performed better than the MRI-only models, which show that additional variables contribute relevant information for prediction of hippocampal atrophy rate. Unfortunately, the analysis of the full models included only about half of the subjects from the MRI-only analysis, because CSF was measured only

Table 7
Variables selected by the feature selection in Analysis 4

L1 + L2 ($\alpha = 0.5$)	L1 ($\alpha = 1.0$)
Volume of left hippocampus	Volume of left hippocampus
Volume of right hippocampus	Volume of right hippocampus
Volume of left amygdala	Volume of left amygdala
Volume of right amygdala	Volume of left inferior lateral ventricle
Volume of left inferior lateral ventricle	Volume of right inferior lateral ventricle
Volume of right inferior lateral ventricle	Volume of left medial orbital gyrus
Volume of left medial orbital gyrus	TBM: right amygdala
TBM: left amygdala	TBM: left cerebellum exterior
TBM: right amygdala	TBM: left hippocampus
TBM: left cerebellum exterior	TBM: right inferior lateral ventricle
TBM: left hippocampus	TBM: right entorhinal area
TBM: right inferior lateral ventricle	TBM: left middle temporal gyrus
TBM: right entorhinal area	TBM: right precuneus
TBM: left middle temporal gyrus	TBM: right parahippocampal gyrus
TBM: right precuneus	VBM: global
TBM: right parahippocampal gyrus	VBM: left hippocampus
VBM: global	VBM: left inferior lateral ventricle
VBM: left hippocampus	VBM: cerebellar vermal lobules VI-VII
VBM: left inferior lateral ventricle	VBM: left inferior temporal gyrus
VBM: cerebellar vermal lobules VI-VII	VBM: right parahippocampal gyrus
VBM: left inferior temporal gyrus	
VBM: left lingual gyrus	

L1 + L2, L1 and L2 norms used in the regularization; L1, L1 norm used in the regularization; α, parameter for controlling weights of L1 and L2 norms; VBM, voxel-based morphometry; TBM, tensor-based morphometry.

in a sub-group of the patients in ADNI 1. Others have predicted progression of AD by predicting conversion from MCI to AD. Nonetheless, in those studies combination of several data modalities provided better prediction performance [5–7, 9]. Ewers et al. [8] found that models with multiple variables performed better than single marker models, but the improvement was not statistically significant. In the study conducted by Gray et al. [11], models with multimodality data performed significantly better than models including only features from CSF, genetic tests, or FDG-PET, but no statistically significant difference was found between the MRI-only model and the combined model.

In all analyses, LAD models performed better than LMS models. This is quite expected because LAD models are known to be less sensitive to outliers than LMS models. This suggests that our data contained some outlying values. The difference between regression with both L1 and L2 norms (elastic net) and regression with only L1 norm (lasso) was only minor in this study. Zou and Hastie [37] have demonstrated that an elastic net often outperformed lasso with simulated and real world data. Lasso is a more stringent regularization technique than elastic net: it selects only one variable from the group of correlating variables and it does not care which one is selected, whereas elastic net can select groups of correlating variables [37].

Variables selected in all cross-validation folds can be considered the most robust variables for the prediction of hippocampal atrophy rate in this study. These variables included information from almost all data modalities: word recall tasks, orientation and total scores from ADAS-cog, CSF amyloid-β and *p*-tau, MRI features describing hippocampus, inferior lateral ventricles, and amygdala. In the case of the MRI-only models, many of the selected brain regions are part of the medial temporal lobe which is known to be affected in AD [44, 45]. Similar variables have been associated with hippocampal atrophy in other studies using different approaches. van de Pol et al. [46] studied a MCI cohort with two years of follow-up. They divided the cohort into tertiles based on annualized hippocampal atrophy rate (absent, moderate, and severe) and found that older age, poorer general cognition measured with ADAS-cog version for MCI, APOE $\varepsilon 4$ prevalence, and hippocampal volumes at baseline were associated with accelerated hippocampal atrophy rate. Gender, episodic memory, executive functioning, vascular risk factors, and MRI features including whole brain volume, white matter hyperintensities, and lacunes did not differ between the three atrophy rate groups. Henneman

et al. [47] used stepwise linear regression to predict hippocampal atrophy rate and they found that CSF *p*-tau levels, baseline memory function measured with Visual Association Test, and visual rating of medial temporal lobe atrophy were associated with hippocampal atrophy rate when corrected for age and gender. Stricker et al. [48] used linear mixed effects model and found also that baseline *p*-tau and amyloid-β were associated with hippocampal atrophy rate over time. It was difficult to compare the performance of our model to the performance of the other models predicting hippocampal atrophy rate, since they did not report RMSE values, correlation coefficients, accuracies, sensitivities, or specificities.

Our models underestimated the real change in hippocampal volume at the higher atrophy rate levels, especially the MRI-only models. We tried to improve the MRI-only models with addition of quadratic terms, but this did not alleviate the problem. Most probably there are other underlying factors that were not included in the current analysis and could explain higher atrophy rates of the subjects. For example, following clinical conditions have been associated with hippocampal atrophy and were not taken into account in our study: cardiovascular disease, cardiac arrest, atrial fibrillation, diabetes, hypertension, obesity, obstructive sleep apnea, vitamin B_{12} deficiency, mood disorders, post-traumatic stress disorder, and head trauma [49]. Inclusion of variables describing aforementioned conditions might improve our models.

The hippocampus is not the only brain region being affected in AD. It is part of the medial temporal lobe, which is affected already at an early phase of the disease. Later in MCI, the disease proceeds to the basal temporal lobe and paralimbic cortical areas, such as the posterior cingulate gyrus and precuneus, and at the onset of dementia, it spreads to the multimodal association neocortices [50]. Even though the disease progresses to different parts of the brain, the hippocampus and amygdala show the fastest atrophy rates at all stages of the disease from normal cognition to MCI and AD [51, 52]. Prediction models aiming to predict disease progression at later phases of the disease could predict, e.g., atrophy rate of the whole brain or posterior cingulate. It would also be interesting to see whether the models developed using our approach for other brain regions would underestimate the change at the higher atrophy rate levels as our current models for hippocampal atrophy did. However, the significance of atrophy in other brain regions is not as well established as is the significance of hippocampal atrophy, e.g., a recent review [53] concluded that hippocampal atrophy is closely related to episodic memory performance and further studies are needed to verify an association between volumes of other brain regions, like posterior cingulate gyrus or precuneus, and memory performance.

The MRI-only model generalized well across cohorts. Interestingly, validation of the MRI-only model with the independent validation data from AIBL study provided higher correlation values between the observed and predicted atrophy rates than cross-validation with ADNI 1. One explanation for this is that the proportion of subjects with MCI or AD was considerably lower in AIBL than in ADNI 1. In addition, we were not able to evaluate the full models with the independent cohort because the AIBL did not include all relevant variables. These are limitations in our study and it would be important to validate the models with another cohort with full set of variables and more subjects with MCI and AD. ADNI 2 might be an appropriate option to try.

Another limitation of our study is that ADNI 1 had follow-up data at months 12 and 24 but AIBL had it only at month 18. Because of the time difference, atrophy rate over 18 months in AIBL was scaled to correspond to atrophy rate over one year in ADNI 1. In scaling, we assumed that the change would be linear. Jack et al. [54] proposed that change in hippocampal volume over time would have a sigmoidal shape. However, the time difference in this study is so small compared to the whole time spectrum of the disease that the linear scaling can be assumed to be valid.

The third limitation in our approach is that hippocampal atrophy is not only related to AD, but it can be affected by other neurodegenerative diseases. In addition, a proportion of healthy or MCI subjects show hippocampal atrophy but they do not develop AD (e.g., suspected non-Alzheimer pathophysiology). However, it would be interesting to see whether our model is able to predict hippocampal atrophy rate in patients with other biomarker data consistent with the AD pathophysiological process, e.g., PET or CSF evidence of cerebral amyloid burden. Thus, this might be extended to preclinical AD, a group of special interest in clinical trials, aiming to study whether an intervention at this early phase of the disease would provide cure or delay progression to more severe stages.

In conclusion, hippocampal atrophy rate at 24 months was predicted using multivariate regularized regression models with data measured at baseline.

The models with cognitive and neuropsychological test results, CSF biomarkers, APOE, and features extracted from MRI performed better than models including only MRI features. The models underestimated the real change at higher atrophy rates. The MRI-only models generalized well across cohorts. However, it would be beneficial to validate the models with other cohorts with full set of variables and including more patients with MCI and AD. These kinds of models can potentially help in selection of suitable patients into clinical trials aiming to test disease-modifying drugs for AD, e.g., by helping to find subjects predicted to have a faster rate of hippocampal atrophy and, thus, higher likelihood of progressing to AD.

ACKNOWLEDGMENTS

This work was carried out in the project European Medical Information Framework (EMIF) which receives support from the Innovative Medicines Initiative Joint Undertaking (IMI-JU) under grant agreement n° 115372, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and the European Federation of Pharmaceutical Industries and Associations' (EFPIA) in kind contribution.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (http://www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Authors' disclosures available online (https://www.j-alz.com/manuscript-disclosures/18-0484r3).

The authors meet criteria for authorship as recommended by the International Committee of Medical Journal Editors (ICMJE). The authors received compensation (i.e., salary) as employees of their respective organization. Any views expressed in this document represent the personal opinions of the authors, and not those of their respective employer. The authors' respective organization was given the opportunity to review the manuscript for medical and scientific accuracy as well as intellectual property considerations.

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: http://dx.doi.org/10.3233/JAD-180484.

REFERENCES

[1] Alzheimer's Association (2016) 2016 Alzheimer's disease facts and figures. *Alzheimers Dement* **12**, 459-509.

[2] Serrano-Pozo A, Frosch MP, Masliah E, Hyman BT (2011) Neuropathological alterations in Alzheimer disease. *Cold Spring Harb Perspect Med* **1**, a006189.

[3] Drago V, Babiloni C, Bartrés-Faz D, Caroli A, Bosch B, Hensch T, Didic M, Klafki H-W, Pievani M, Jovicich J, Venturi L, Spitzer P, Vecchio F, Schoenknecht P, Wiltfang J, Redolfi A, Forloni G, Blin O, Irving E, Davis C, Hårdemark H, Frisoni GB (2011) Disease tracking markers for Alzheimer's disease at the prodromal (MCI) stage. *J Alzheimers Dis* **26**, 159-199.

[4] Barnes J, Bartlett JW, van de Pol LA, Loy CT, Scahill RI, Frost C, Thompson P, Fox NC (2009) A meta-analysis of hippocampal atrophy rates in Alzheimer's disease. *Neurobiol Aging* **30**, 1711-1723.

[5] Devanand D, Pradhaban G, Liu X, Khandji A, De Santi S, Segal S, Rusinek H, Pelton G, Honig L, Mayeux R, Stern Y, Tabert M, de Leon M (2007) Hippocampal and entorhinal atrophy in mild cognitive impairment: Prediction of Alzheimer disease. *Neurology* **68**, 828-836.

[6] Devanand DP, Liu X, Tabert MH, Pradhaban G, Cuasay K, Bell K, de Leon MJ, Doty RL, Stern Y, Pelton GH (2008) Combining early markers strongly predicts conversion from mild cognitive impairment to Alzheimer's disease. *Biol Psychiatry* **64**, 871-879.

[7] Mattila J, Koikkalainen J, Virkki A, Simonsen A, van Gils M, Waldemar G, Soininen H, Lötjönen J, the Alzheimer's Disease Neuroimaging Initiative (2011) A disease state fingerprint for evaluation of Alzheimer's disease. *J Alzheimers Dis* **27**, 163-176.

[8] Ewers M, Walsh C, Trojanowski JQ, Shaw LM, Petersen RC, Jack CR, Feldman HH, Bokde ALW, Alexander GE, Scheltens P, Vellas B, Dubois B, Weiner M, Hampel H, North American Alzheimer's Disease Neuroimaging Initiative (ADNI) (2012) Prediction of conversion from mild cognitive impairment to Alzheimer's disease dementia based upon biomarkers and neuropsychological test performance. *Neurobiol Aging* **33**, 1203-1214.

[9] Cui Y, Liu B, Luo S, Zhen X, Fan M, Liu T, Zhu W, Park M, Jiang T, Jin JS, Alzheimer's Disease Neuroimaging Initiative (2011) Identification of conversion from mild cognitive impairment to Alzheimer's disease using multivariate predictors. *PLoS One* **6**, e21896.

[10] Zhang D, Shen D, the Alzheimer's Disease Neuroimaging Initiative (2012) Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. *PLoS One* **7**, e33182.

[11] Gray KR, Aljabar P, Heckemann RA, Hammers A, Rueckert D, Alzheimer's Disease Neuroimaging Initiative (2013) Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *Neuroimage* **65**, 167-175.

[12] Llano DA, Laforet G, Devanarayan V, the Alzheimer's Disease Neuroimaging Initiative (2011) Derivation of a new ADAS-cog composite using tree-based multivariate analysis: Prediction of conversion from mild cognitive impairment to Alzheimer disease. *Alzheimer Dis Assoc Disord* **25**, 73-84.

[13] Eskildsen SF, Coupé P, Fonov VS, Pruessner JC, Collins DL, the Alzheimer's Disease Neuroimaging Initiative (2015) Structural imaging biomarkers of Alzheimer's disease: Predicting disease progression. *Neurobiol Aging* **36**, S23-S31.

[14] Zhang D, Shen D, the Alzheimer's Disease Neuroimaging Initiative (2012) Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. *Neuroimage* **59**, 895-907.

[15] Zhou J, Liu J, Narayan VA, Ye J, the Alzheimer's Disease Neuroimaging Initiative (2013) Modeling disease progression via multi-task learning. *Neuroimage* **78**, 233-248.

[16] Kovacevic S, Rafii MS, Brewer JB, the Alzheimer's Disease Neuroimaging Initiative (2009) High-throughput, fully automated volumetry for prediction of MMSE and CDR decline in mild cognitive impairment. *Alzheimer Dis Assoc Disord* **23**, 139-145.

[17] Allen GI, Amoroso N, Anghel C, Balagurusamy V, Bare CJ, Beaton D, Bellotti R, Bennett DA, Boehme KL, Boutros PC, Caberlotto L, Caloian C, Campbell F, Chaibub Neto E, Chang Y-C, Chen B, Chen C-Y, Chien T-Y, Clark T, Das S, Davatzikos C, Deng J, Dillenberger D, Dobson RJB, Dong Q, Doshi J, Duma D, Errico R, Erus G, Everett E, Fardo DW, Friend SH, Fröhlich H, Gan J, St George-Hyslop P, Ghosh SS, Glaab E, Green RC, Guan Y, Hong M-Y, Huang C, Hwang J, Ibrahim J, Inglese P, Iyappan A, Jiang Q, Katsumata Y, Kauwe JSK, Klein A, Kong D, Krause R, Lalonde E, Lauria M, Lee E, Lin X, Liu Z, Livingstone J, Logsdon BA, Lovestone S, Ma T, Malhotra A, Mangravite LM, Maxwell TJ, Merrill E, Nagorski J, Namasivayam A, Narayan M, Naz M, Newhouse SJ, Norman TC, Nurtdinov RN, Oyang Y-J, Pawitan Y, Peng S, Peters MA, Piccolo SR, Praveen P, Priami C, Sabelnykova VY, Senger P, Shen X, Simmons A, Sotiras A, Stolovitzky G, Tangaro S, Tateo A, Tung Y-A, Tustison NJ, Varol E, Vradenburg G, Weiner MW, Xiao G, Xie L, Xie Y, Xu J, Yang H, Zhan X, Zhou Y, Zhu F, Zhu H, Zhu S (2016) Crowdsourced estimation of cognitive decline and resilience in Alzheimer's disease. *Alzheimers Dement* **12**, 645-653.

[18] Zhu F, Panwar B, Dodge HH, Li H, Hampstead BM, Albin RL, Paulson HL, Guan Y (2016) COMPASS: A computational model to predict changes in MMSE scores 24-months after initial assessment of Alzheimer's disease. *Sci Rep* **6**, 34567.

[19] Huang L, Jin Y, Gao Y, Thung K-H, Shen D, the Alzheimer's Disease Neuroimaging Initiative (2016) Longitudinal clinical score prediction in Alzheimer's disease with soft-split sparse regression based random forest. *Neurobiol Aging* **46**, 180-191.

[20] Samtani MN, Raghavan N, Novak G, Nandy P, Narayan VA, the Alzheimer's Disease Neuroimaging Initiative (2014) Disease progression model for Clinical Dementia Rating-Sum of Boxes in mild cognitive impairment and Alzheimer's subjects from the Alzheimer's Disease Neuroimaging Initiative. *Neuropsychiatr Dis Treat* **10**, 929-952.

[21] Jack CR, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, Shaw LM, Vemuri P, Wiste HJ, Weigand SD, Lesnick TG, Pankratz VS, Donohue MC, Trojanowski JQ (2013) Tracking pathophysiological processes in Alzheimer's disease: An updated hypothetical model of dynamic biomarkers. *Lancet Neurol* **12**, 207-216.

[22] Hampel H, Bürger K, Teipel SJ, Bokde ALW, Zetterberg H, Blennow K (2008) Core candidate neurochemical and imaging biomarkers of Alzheimer's disease. *Alzheimers Dement* **4**, 38-48.

[23] Hampel H, Frank R, Broich K, Teipel SJ, Katz RG, Hardy J, Herholz K, Bokde ALW, Jessen F, Hoessler YC, Sanhai WR, Zetterberg H, Woodcock J, Blennow K (2010) Biomarkers for Alzheimer's disease: Academic, industry and regulatory perspectives. *Nat Rev Drug Discov* **9**, 560-574.

[24] European Medicines Agency (2011) Qualification opinion of low hippocampal volume (atrophy) by MRI for use in clinical trials for regulatory purpose - in pre-dementia stage of Alzheimer's disease, https://www.ema.europa.eu/documents/regulatory-procedural-guideline/qualification-opinion-low-hippocampal-volume-atrophy-magnetic-resonance-imaging-use-clinical-trials_en.pdf, Accessed 25 January 2019.

[25] Food and Drug Administration (2015) Biomarker Letter of Support, https://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/DrugDevelopmentToolsQualificationProgram/BiomarkerQualificationProgram/UCM605354.pdf, Accessed 25 January 2019.

[26] Alzheimer's Disease Neuroimaging Initiative (2006) ADNI 1 Procedures Manual, http://adni.loni.usc.edu/wp-content/uploads/2010/09/ADNI_GeneralProceduresManual.pdf, Accessed 12 June 2018.

[27] Ellis KA, Bush AI, Darby D, De Fazio D, Foster J, Hudson P, Lautenschlager NT, Lenzo N, Martins RN, Maruff P, Masters C, Milner A, Pike K, Rowe C, Savage G, Szoeke C, Taddei K, Villemagne V, Woodward M, Ames D, AIBL Research Group (2009) The Australian Imaging, Biomark-

ers and Lifestyle (AIBL) study of aging: Methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. *Int Psychogeriatr* **21**, 672-687.

[28] Shaw LM, Vanderstichele H, Knapik-Czajka M, Clark CM, Aisen PS, Petersen RC, Blennow K, Soares H, Simon A, Lewczuk P, Dean R, Siemers E, Potter W, Lee VM-Y, Trojanowski JQ, Alzheimer's Disease Neuroimaging Initiative (2009) Cerebrospinal fluid biomarker signature in Alzheimer's disease neuroimaging initiative subjects. *Ann Neurol* **65**, 403-413.

[29] Villemagne VL, Burnham S, Bourgeat P, Brown B, Ellis KA, Salvado O, Szoeke C, Macaulay SL, Martins R, Maruff P, Ames D, Rowe CC, Masters CL (2013) Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: A prospective cohort study. *Lancet Neurol* **12**, 357-367.

[30] Rowe CC, Ellis KA, Rimajova M, Bourgeat P, Pike KE, Jones G, Fripp J, Tochon-Danguy H, Morandeau L, O'Keefe G, Price R, Raniga P, Robins P, Acosta O, Lenzo N, Szoeke C, Salvado O, Head R, Martins R, Masters CL, Ames D, Villemagne VL (2010) Amyloid imaging results from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging. *Neurobiol Aging* **31**, 1275-1283.

[31] Koikkalainen J, Rhodius-Meester H, Tolonen A, Barkhof F, Tijms B, Lemstra AW, Tong T, Guerrero R, Schuh A, Ledig C, Rueckert D, Soininen H, Remes AM, Waldemar G, Hasselbalch S, Mecocci P, van der Flier W, Lötjönen J (2016) Differential diagnosis of neurodegenerative diseases using structural MRI data. *Neuroimage Clin* **11**, 435-449.

[32] Lötjönen JM, Wolz R, Koikkalainen JR, Thurfjell L, Waldemar G, Soininen H, Rueckert D, the Alzheimer's Disease Neuroimaging Initiative (2010) Fast and robust multi-atlas segmentation of brain magnetic resonance images. *Neuroimage* **49**, 2352-2365.

[33] Koikkalainen J, Lötjönen J, Thurfjell L, Rueckert D, Waldemar G, Soininen H, the Alzheimer's Disease Neuroimaging Initiative (2011) Multi-template tensor-based morphometry: Application to analysis of Alzheimer's disease. *Neuroimage* **56**, 1134-1144.

[34] Ashburner J, Friston KJ (2000) Voxel-based morphometry—the methods. *Neuroimage* **11**, 805-821.

[35] Lötjönen J, Ledig C, Koikkalainen J, Wolz R, Thurfjell L, Soininen H, Ourselin S, Rueckert D, the Alzheimer's Disease Neuroimaging Initiative (2014) Extended boundary shift integral. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)* IEEE, Beijing, pp. 854-857.

[36] Freeborough PA, Fox NC (1997) The boundary shift integral: An accurate and robust measure of cerebral volume changes from registered repeat MRI. *IEEE Trans Med Imaging* **16**, 623-629.

[37] Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. *J R Stat Soc Ser B Stat Methodol* **67**, 301-320.

[38] Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* **33**, 1-22.

[39] Dasgupta M, Mishra SK (2004) *Least absolute deviation estimation of linear econometric models: A literature review*, University Library of Munich, Germany.

[40] Beckett LA, Harvey DJ, Gamst A, Donohue M, Kornak J, Zhang H, Kuo JH, Alzheimer's Disease Neuroimaging Initiative (2010) The Alzheimer's Disease Neuroimaging Initiative: Annual change in biomarkers and clinical outcomes. *Alzheimers Dement* **6**, 257-264.

[41] Cronk BB, Johnson DK, Burns JM, Alzheimer's Disease Neuroimaging Initiative (2010) Body mass index and cognitive decline in mild cognitive impairment. *Alzheimer Dis Assoc Disord* **24**, 126-130.

[42] Fjell AM, Walhovd KB, Fennema-Notestine C, McEvoy LK, Hagler DJ, Holland D, Brewer JB, Dale AM (2009) One-year brain atrophy evident in healthy aging. *J Neurosci* **29**, 15223-15231.

[43] Koikkalainen J, Pölönen H, Mattila J, van Gils M, Soininen H, Lötjönen J, the Alzheimer's Disease Neuroimaging Initiative (2012) Improved classification of Alzheimer's disease data via removal of nuisance variability. *PLoS One* **7**, e31112.

[44] Frisoni GB, Fox NC, Jack CR, Scheltens P, Thompson PM (2010) The clinical use of structural MRI in Alzheimer disease. *Nat Rev Neurol* **6**, 67-77.

[45] Fox NC, Schott JM (2004) Imaging cerebral atrophy: Normal ageing to Alzheimer's disease. *Lancet* **363**, 392-394.

[46] van de Pol LA, van der Flier WM, Korf ESC, Fox NC, Barkhof F, Scheltens P (2007) Baseline predictors of rates of hippocampal atrophy in mild cognitive impairment. *Neurology* **69**, 1491-1497.

[47] Henneman WJP, Vrenken H, Barnes J, Sluimer IC, Verwey NA, Blankenstein MA, Klein M, Fox NC, Scheltens P, Barkhof F, van der Flier WM (2009) Baseline CSF p-tau levels independently predict progression of hippocampal atrophy in Alzheimer disease. *Neurology* **73**, 935-940.

[48] Stricker NH, Dodge HH, Dowling NM, Han SD, Erosheva EA, Jagust WJ, the Alzheimer's Disease Neuroimaging Initiative (2012) CSF biomarker associations with change in hippocampal volume and precuneus thickness: Implications for the Alzheimer's pathological cascade. *Brain Imaging Behav* **6**, 599-609.

[49] Fotuhi M, Do D, Jack C (2012) Modifiable factors that alter the size of the hippocampus with ageing. *Nat Rev Neurol* **8**, 189-202.

[50] Vemuri P, Jack CR (2010) Role of structural MRI in Alzheimer's disease. *Alzheimers Res Ther* **2**, 23.

[51] McDonald CR, McEvoy LK, Gharapetian L, Fennema-Notestine C, Hagler DJ, Holland D, Koyama A, Brewer JB, Dale AM, Alzheimer's Disease Neuroimaging Initiative (2009) Regional rates of neocortical atrophy from normal aging to early Alzheimer disease. *Neurology* **73**, 457-465.

[52] Fjell AM, Walhovd KB, Fennema-Notestine C, McEvoy LK, Hagler DJ, Holland D, Brewer JB, Dale AM, Alzheimer's Disease Neuroimaging Initiative Neuroimaging (2010) CSF biomarkers in prediction of cerebral and clinical change in mild cognitive impairment and Alzheimer's disease. *J Neurosci* **30**, 2088-2101.

[53] Bayram E, Caldwell JZK, Banks SJ (2018) Current understanding of magnetic resonance imaging biomarkers and memory in Alzheimer's disease. *Alzheimers Dement (N Y)* **4**, 395-413.

[54] Jack CR, Knopman DS, Jagust WJ, Shaw LM, Aisen PS, Weiner MW, Petersen RC, Trojanowski JQ (2010) Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol* **9**, 119-128.

SUPPLEMENTARY MATERIAL

Supplementary Table 1 lists all variables used as predictors. Analysis 1 included all variables and Analysis 2, Analysis 3 and Analysis 4 included only MRI features. Supplementary Table 2 presents demographics of the study populations separately for each diagnostic class. Supplementary Table 3 shows absolute loss of hippocampal volume between the baseline and 24-month follow-up. Supplementary Figure 1 presents the observed and predicted hippocampal atrophy rate for the different models when only L1 norm was used in the regularization.

Supplementary Table 1. Predictor variables

Cognitive and neuropsychological	**Items**	**N**
Mini-Mental State Examination	All items + total score	31
Alzheimer's Disease Assessment Scale – cognitive subscale	All items + modified total score	14
Clinical Dementia Rating	All items + global score + total score	8
Clock draw	All items + total score	6
Clock Copy	All items + total score	6
Auditory Verbal Learning Test	Trials I-VI + List B (total score, intrusions)	14
Digit Span	Forward + Backward (total score, length)	4
Category Fluency Test	Animals + Vegetables (total score, perseverations, intrusions)	6
Trail Making Test	Part A + Part B	6
Digit Symbol Substitution	All items	1
Boston Naming Test	All items	6
Auditory Verbal Learning Test Delayed Recall	Delayed recall + Recognition	4
American National Adult Reading Test	All items	1
Cerebrospinal fluid sample		
Amyloid-β		
Tau		
P-tau		
Genetic		
APOE: type of allele 1 (ε2, ε3, or ε4) as dummy coded variables		
APOE: type of allele 2 (ε2, ε3, or ε4) as dummy coded variables		
MRI: volumetry of hippocampus using an atlas from ADNI [32]		

Right Hippocampus
Left Hippocampus

MRI: volumetry, tensor-based morphometry, and voxel-based morphometry for the following brain regions from the Neuromorphometrics atlas	
Global	Right Lingual Gyrus
3rd Ventricle	Left Lingual Gyrus
4th Ventricle	Right Lateral Orbital Gyrus
5th Ventricle	Left Lateral Orbital Gyrus
Right Accumbens Area	Right Middle Cingulate Gyrus
Left Accumbens Area	Left Middle Cingulate Gyrus
Right Amygdala	Right Medial Frontal Cortex
Left Amygdala	Left Medial Frontal Cortex
Brain Stem	Right Middle Frontal Gyrus
Right Caudate	Left Middle Frontal Gyrus
Left Caudate	Right Middle Occipital Gyrus
Right Cerebellum Exterior	Left Middle Occipital Gyrus
Left Cerebellum Exterior	Right Medial Orbital Gyrus
Right Cerebellum White Matter	Left Medial Orbital Gyrus
Left Cerebellum White Matter	Right Postcentral Gyrus Medial Segment
Right Cerebral Exterior	Left Postcentral Gyrus Medial Segment
Left Cerebral Exterior	Right Precentral Gyrus Medial Segment
Right Cerebral White Matter	Left Precentral Gyrus Medial Segment
Left Cerebral White Matter	Right Superior Frontal Gyrus Medial Segment
CSF	Left Superior Frontal Gyrus Medial Segment
Right Hippocampus	Right Middle Temporal Gyrus
Left Hippocampus	Left Middle Temporal Gyrus
Right Inferior Lateral Ventricle	Right Occipital Pole
Left Inferior Lateral Ventricle	Left Occipital Pole
Right Lateral Ventricle	Right Occipital Fusiform Gyrus
Left Lateral Ventricle	Left Occipital Fusiform Gyrus
Right Pallidum	Right Opercular Part of the Inferior Frontal Gyrus
Left Pallidum	Left Opercular Part of the Inferior Frontal Gyrus
Right Putamen	Right Orbital Part of the Inferior Frontal Gyrus
Left Putamen	Left Orbital Part of the Inferior Frontal Gyrus
Right Thalamus Proper	Right Posterior Cingulate Gyrus
Left Thalamus Proper	Left Posterior Cingulate Gyrus
Right Ventral DC	Right Precuneus
Left Ventral DC	Left Precuneus
Right Vessel	Right Parahippocampal Gyrus
Left Vessel	Left Parahippocampal Gyrus
Optic Chiasm	Right Posterior Insula

Cerebellar Vermal Lobules|I-V	Left Posterior Insula
Cerebellar Vermal Lobules|VI-VII	Right Parietal Operculum
Cerebellar Vermal Lobules|VIII-X	Left Parietal Operculum
Right Basal Forebrain	Right Postcentral Gyrus
Left Basal Forebrain	Left Postcentral Gyrus
Right Anterior Cingulate Gyrus	Right Posterior Orbital Gyrus
Left Anterior Cingulate Gyrus	Left Posterior Orbital Gyrus
Right Anterior Insula	Right Planum Polare
Left Anterior Insula	Left Planum Polare
Right Anterior Orbital Gyrus	Right Precentral Gyrus
Left Anterior Orbital Gyrus	Left Precentral Gyrus
Right Angular Gyrus	Right Planum Temporale
Left Angular Gyrus	Left Planum Temporale
Right Calcarine Cortex	Right Subcallosal Area
Left Calcarine Cortex	Left Subcallosal Area
Right Central Operculum	Right Superior Frontal Gyrus
Left Central Operculum	Left Superior Frontal Gyrus
Right Cuneus	Right Supplementary Motor Cortex
Left Cuneus	Left Supplementary Motor Cortex
Right Entorhinal Area	Right Supramarginal Gyrus
Left Entorhinal Area	Left Supramarginal Gyrus
Right Frontal Operculum	Right Superior Occipital Gyrus
Left Frontal Operculum	Left Superior Occipital Gyrus
Right Frontal Pole	Right Superior Parietal Lobule
Left Frontal Pole	Left Superior Parietal Lobule|
Right Fusiform Gyrus	Right Superior Temporal Gyrus
Left Fusiform Gyrus	Left Superior Temporal Gyrus
Right Gyrus Rectus	Right Temporal Pole
Left Gyrus Rectus	Left Temporal Pole
Right Inferior Occipital Gyrus	Right Triangular Part of the Inferior Frontal Gyrus
Left Inferior Occipital Gyrus	Left Triangular Part of the Inferior Frontal Gyrus
Right Inferior Temporal Gyrus	Right Transverse Temporal Gyrus
Left Inferior Temporal Gyrus	Left Transverse Temporal Gyrus

Supplementary Table 2. Demographics of the study population at the baseline, stratified by the diagnostic class

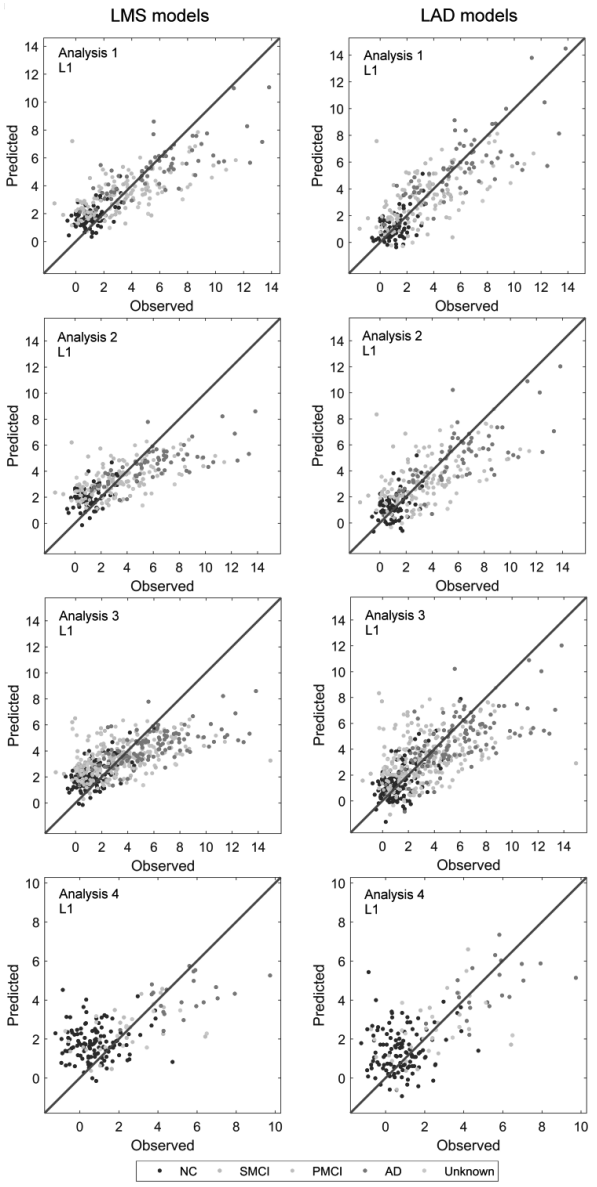
	ADNI 1										AIBL				
	Analysis 1 & 2					Analysis 3					Analysis 4				
	NC	SMCI	PMCI	AD	Unknown	NC	SMCI	PMCI	AD	Unknown	NC	SMCI	PMCI	AD	Unknown
N	75 (27)	61 (22)	64 (23)	63 (22)	18 (6)	151 (28)	118 (22)	118 (22)	109 (21)	34 (6)	119 (68)	12 (7)	6 (3)	21 (12)	18 (10)
Age (years)	76.0 ± 4.7	74.4 ± 7.2	75.3 ± 7.3	75.1 ± 7.6	73.7 ± 7.4	76.1 ± 4.8	75.0 ± 7.0	74.9 ± 6.8	75.6 ± 7.4	75.1 ± 6.9	71.6 ± 7.0	76.8 ± 6.1	76.3 ± 6.7	70.6 ± 8.1	70.2 ± 6.4
Gender															
Male	34 (45)	42 (69)	42 (66)	35 (56)	14 (78)	76 (50)	78 (66)	75 (64)	56 (51)	24 (71)	58 (49)	6 (50)	4 (67)	8 (38)	12 (67)
Female	41 (55)	19 (31)	22 (34)	28 (44)	4 (22)	75 (50)	40 (34)	43 (36)	53 (49)	10 (29)	61 (51)	6 (50)	2 (33)	13 (62)	6 (33)
MMSE	29.2 ± 1.0	27.2 ± 1.6	26.6 ± 1.9	23.4 ± 1.9	28.4 ± 1.4	29.2 ± 0.9	27.5 ± 1.7	26.7 ± 1.6	23.2 ± 2.0	28.4 ± 1.6	28.9 ± 1.2	27.8 ± 1.7	26.7 ± 2.2	22.3 ± 4.6	27.6 ± 1.4
HC (mm^3)	4283 ± 517	3864 ± 468	3561 ± 587	3405 ± 684	4205 ± 477	4250 ± 559	3888 ± 533	3609 ± 640	3377 ± 657	4096 ± 506	4299 ± 517	4134 ± 647	3638 ± 603	3678 ± 642	4238 ± 501
Aβ	207 ± 54	167 ± 58	145 ± 39	143 ± 41	193 ± 62	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Aβ+	27 (36)	43 (70)	58 (91)	57 (90)	8 (44)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

NC: healthy elderly control; SMCI: stable mild cognitive impairment; PMCI: progressive mild cognitive impairment; AD: Alzheimer's disease; Unknown: diagnosis changed otherwise than from MCI to AD; N: number of subjects; MMSE: total score of Mini-Mental State Examination; HC: hippocampal volume (left + right); Aβ: amyloid-β; Aβ+; number of amyloid positive subjects, i.e. subjects whose amyloid-β was below 192 pg/ml [28]. NA: not available due to lack of cerebrospinal fluid samples.Values are expressed as mean ± standard deviation or as count (percentage).

Supplementary Table 3. Absolute loss of hippocampal volume [mm^3] between the baseline and 24-month follow-up

	Analysis 1 & 2	Analysis 3	Analysis 4
NC	88 ± 71	95 ± 84	64 ± 89
SMCI	190 ± 144	181 ± 163	186 ± 143
PMCI	316 ± 178	305 ± 161	214 ± 98
AD	384 ± 197	369 ± 183	380 ± 138

NC: healthy elderly control; SMCI: stable mild cognitive impairment; PMCI: progressive mild cognitive impairment; AD: Alzheimer's disease; Unknown: diagnosis changed otherwise than from MCI to AD.



Supplementary Figure 1. Observed and predicted annual hippocampal atrophy rate for the models with L1 norms in the regularization. LMS: least mean square regression; LAD: least absolute deviation regression; NC: normal control; SMCI: stable mild cognitive impairment; PMCI: progressive mild cognitive impairment; AD: Alzheimer's disease; Unknown: diagnosis of the subject changed otherwise than from MCI to AD.

PUBLICATION IV

Disease-related determinants are associated with mortality in dementia due to Alzheimer's disease

Rhodius-Meester H. F. M., Liedes H., Koene T., Lemstra A. W., Teunissen C. E., Barkhof F., Scheltens P., van Gils M., Lötjönen J. & van der Flier W. M.

Alzheimer's Research & Therapy, 2018, 10(23)
https://doi.org/10.1186/s13195-018-0348-0

Rhodius-Meester et al. Alzheimer's Research & Therapy (2018) 10:23
https://doi.org/10.1186/s13195-018-0348-0

Alzheimer's
Research & Therapy

RESEARCH Open Access

Disease-related determinants are associated with mortality in dementia due to Alzheimer's disease

Hanneke F. M. Rhodius-Meester[1*], Hilkka Liedes[2], Ted Koene[3], Afina W. Lemstra[1], Charlotte E. Teunissen[4], Frederik Barkhof[5,6,7], Philip Scheltens[1], Mark van Gils[2], Jyrki Lötjönen[2,8] and Wiesje M. van der Flier[1,9]

Abstract

Background: Survival after dementia diagnosis varies considerably. Previous studies were focused mainly on factors related to demographics and comorbidity rather than on Alzheimer's disease (AD)-related determinants. We set out to answer the question whether markers with proven diagnostic value also have prognostic value. We aimed to identify disease-related determinants associated with mortality in patients with AD.

Methods: We included 616 patients (50% female; age 67 ± 8 years; mean Mini Mental State Examination score 22 ± 3) with dementia due to AD from the Amsterdam Dementia Cohort. Information on mortality was obtained from the Dutch Municipal Register. We used age- and sex-adjusted Cox proportional hazards analysis to study associations of baseline demographics, comorbidity, neuropsychology, magnetic resonance imaging (MRI) (medial temporal lobe, global cortical and parietal atrophy, and measures of small vessel disease), and cerebrospinal fluid (CSF) (β-amyloid 1–42, total tau, and tau phosphorylated at threonine 181 [p-tau]) with mortality (outcome). In addition, we built a multivariate model using forward selection.

Results: After an average of 4.9 ± 2.0 years, 213 (35%) patients had died. Age- and sex-adjusted Cox models showed that older age (HR 1.29 [95% CI 1.12–1.48]), male sex (HR 1.60 [95% CI 1.22–2.11]), worse scores on cognitive functioning (HR 1.14 [95% CI 1.01-1.30] to 1.31 [95% CI 1.13–1.52]), and more global and hippocampal atrophy on MRI (HR 1.18 [95% CI 1.01-1.37] and HR 1.18 [95% CI 1.02-1.37]) were associated with increased risk of mortality. There were no associations with comorbidity, level of activities of daily living, apolipoprotein E (*APOE*) ε4 status, or duration of disease. Using forward selection, the multivariate model included a panel of age, sex, cognitive tests, atrophy of the medial temporal lobe, and CSF p-tau.

Conclusions: In this relatively young sample of patients with AD, disease-related determinants were associated with an increased risk of mortality, whereas neither comorbidity nor *APOE* genotype had any prognostic value.

Keywords: Alzheimer's disease, Prognosis, Mortality, Diagnostic test assessment

Background

Dementia due to Alzheimer's disease (AD) is, by definition, a progressive disorder [1]. For patients, a diagnosis is not the endpoint, but rather the beginning of the subsequent trajectory of disease. Physicians are fairly good at establishing an accurate diagnosis, but they are hardly able to predict the course of the disease for the individual patient. In general, patients with dementia due to AD have a shorter life expectancy than the general population, with average survival being between 5 and 10 years [2–4]. Yet, survival time varies considerably between individuals.

Determinants of mortality in AD have been examined in various studies. Most have been focused on demographic factors or on clinical factors such as severity of cognitive impairment, dependency, and comorbidity [5–9]. Male sex and older age have been associated with increased risk of mortality in AD [2, 3, 10]. Cardiovascular diseases and

* Correspondence: h.rhodius@vumc.nl
[1]Alzheimer Center, Department of Neurology, VU University Medical Center, Amsterdam Neuroscience, P.O. Box 7057, 1007, MB, Amsterdam, The Netherlands
Full list of author information is available at the end of the article

risk factors such as hypertension and diabetes mellitus have been identified as determinants of mortality, but only in studies of older patients with dementia [11–15]. Few studies have been focused on AD-specific factors, such as cerebrospinal fluid (CSF) and magnetic resonance imaging (MRI) markers. More severe neuronal degeneration, as reflected by a high total tau (tau) concentration and whole-brain atrophy, has been suggested as a determinant of mortality [12, 16–18]. In one study, microbleeds were associated with mortality in AD, and white matter hyperintensities (WMH) were associated with mortality in all-cause dementia [12].

Researchers in previous studies tended to evaluate only a few prognostic factors per study and included mainly patients aged 75 years and older, who are at risk of mortality owing to their advanced age even without a dementia diagnosis [3, 19]. Prognostic factors may be different for patients with early-onset AD, who are younger and have less comorbidity but are prone to a more aggressive disease course [10, 20–22]. We aimed to investigate the prognostic value of baseline clinical data, including demographics, comorbidity, neuropsychology, and CSF and MRI biomarkers, as determinants of mortality in dementia due to AD.

Methods

Patients

In this longitudinal study, we included 616 patients with a baseline diagnosis of dementia due to AD from a memory clinic-based cohort (the Amsterdam Dementia Cohort) who had a baseline visit between 2000 and 2014 [23]. Subjects were selected if a neuropsychological test battery was available at baseline, with a baseline Mini Mental State Examination (MMSE) score ≥ 16 and a minimum follow-up of 2 years. At baseline, patients received a standardized and multidisciplinary workup, including medical history; physical, neurological, and neuropsychological examinations; MRI; laboratory tests; and lumbar puncture for CSF measurements. Years of education and self-reported duration of complaints were recorded. For the assessment of activities of daily living, we used the Disability Assessment for Dementia (DAD) [24]. We included all data that were collected within 6 months of baseline diagnosis. Diagnoses were made in a multidisciplinary consensus meeting. Patients were diagnosed with probable AD using the criteria of the National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer's Disease and Related Disorders Association; all patients also met the core clinical criteria of the National Institute on Aging-Alzheimer's Association for AD dementia [25, 26].

Medical history

We recorded and defined the presence (yes/no) of hypertension (history of hypertension and/or use of antihypertensive drugs), hypercholesterolemia (history of hypercholesterolemia and/or use of cholesterol-lowering drugs), diabetes (history of diabetes mellitus and/or use of antidiabetic drugs), and cardiovascular disease (at least one of the following: history of coronary heart disease, heart failure, heart disease, peripheral vascular disease, stroke, and/or transient ischemic attack). Furthermore, we dichotomized smoking status (never smoked versus current or history of smoking) and counted the medications used per patient.

Neuropsychological tests

Cognitive function was assessed at baseline with a standardized test battery in which the MMSE was used for global cognitive functioning [27]. For memory, the Visual Association Test (VAT) and the Rey Auditory Verbal Learning Task (RAVLT) were included [28, 29]. To measure mental speed and attention, we used Trail Making Test A (TMT-A) and the forward condition of the digit span. Trail Making Test B (TMT-B) and the backward condition of the digit span were used for executive functioning [30, 31]. Language and executive functioning were tested by category fluency (animals) [32]. Missing data ranged from n = 19 (3%) (digit span forward) to n = 67 (11%) (RAVLT, delayed recall).

MRI

Subjects were scanned as part of clinical workup using a standardized protocol on a 1.0-, 1.5-, or 3.0-T system. All scans were visually rated by trained raters and subsequently evaluated in a consensus meeting with an experienced neuroradiologist [23]. Visual rating of medial temporal lobe atrophy (MTA) was performed using coronal T1-weighted images on a 5-point (0–4) scale from the average score of the left and right sides [33]. Global cortical atrophy (GCA) was assessed visually on axial fluid-attenuated inversion recovery (FLAIR) images (range of scores 0–3) [34]. Parietal atrophy was rated using T1-weighted and FLAIR weighted images viewed in sagittal, axial, and coronal planes by computing an average score of the left and right sides (range 0–3) [35]. WMH were rated on axial FLAIR images using a four-step scale (range 0–3) [36]. Lacunes were defined as deep lesions (3–15 mm) with CSF-like signals on all sequences and were dichotomized as present or absent. Microbleeds were defined as small, round foci of hypointense signal up to 10 mm in brain parenchyma on T2*-weighted gradient echo images. The total number of microbleeds was counted and divided into three categories: zero, one or two, and three or more microbleeds. MRI data were available for 485 (79%) subjects.

CSF

CSF analyses were performed at the Neurochemistry Laboratory at the Department of Clinical Chemistry of the VUmc. CSF was obtained by lumbar puncture between the L3-L4 or L4-L5 intervertebral space by using a 25-gauge needle and collected into polypropylene tubes. Within 2 h, the CSF was centrifuged at 1800 × *g* for 10 minutes at 4 °C, transferred to new polypropylene tubes, and stored at −20 °C until biomarker analysis (within 2 months). β-Amyloid 1–42 ($A\beta_{42}$), tau, and tau phosphorylated at threonine 181 (p-tau) were measured with commercially available enzyme-linked immunosorbent assays (Innotest; Fujirebio, Ghent, Belgium) [37]. CSF data were available for 466 (76%) subjects.

APOE genotyping

DNA was isolated from 10 ml of ethylenediaminetetraacetic acid blood. Apolipoprotein E (*APOE*) genotype was determined using the LightCycler *APOE* mutation detection method (Roche Diagnostics GmbH, Mannheim, Germany). According to *APOE* ε4 allele status, patients were dichotomized into carriers (hetero- and homozygous) and noncarriers. *APOE* status was available for 562 (91%) subjects.

Outcome measure

For each patient, we obtained information on all-cause mortality (died yes/no with a date of death) from the Dutch municipal population register. This register was searched on 19 October 2016. Causes of death cannot be determined from this municipal registry. We defined follow-up duration as the time between the date of baseline AD diagnosis and the date of death or, if alive, between the date of baseline AD diagnosis and 19 October 2016.

Statistical analyses

Statistical analyses were performed using IBM SPSS Statistics version 22 software (IBM, Armonk, NY, USA). $p < 0.05$ was considered significant. Baseline characteristics were compared using parametric and nonparametric tests when appropriate. We used pattern analysis to explore the amount and randomness of missing data. Because missing data were at random, but not completely at random, we imputed all missing data imputed using multiple imputation, in which the missing values were estimated on the basis of other available baseline variables in 15 imputation cycles.

To allow comparison of results on different tests within patients, all continuous variables were standardized to z-scores. All neuropsychological tests, except TMT-A and TMT-B, as well as CSF $A\beta_{42}$, were inverted by computing −1 × z-score, with the result being that a higher score implied more advanced disease. We used Cox proportional hazards models to assess associations between all baseline determinants and mortality, taking into account time to death, using the pooled results of the 15 imputations. Each measure was assessed unadjusted (model 1), adjusted for age and sex (model 2), and adjusted additionally for MMSE and duration of

Table 1 Baseline characteristics of patients with Alzheimer's disease according to outcome

	No. of patients	Alive (*n* = 403)	Died (*n* = 213)	*p* Value
Demographics				
Female sex, *n* (%)	616	218 (54)	91 (43)	0.007
Age, years	616	66 ± 7	69 ± 9	0.000
APOE ε4 carrier, *n* (%)	562	250 (67)	119 (63)	0.280
Years of education	616	11 ± 3	11 ± 3	0.675
Years of complaints	611	3.2 ± 2.6	2.8 ± 2.0	0.066
Years to outcome	616	5.3 ± 1.8	4.3 ± 2.1	0.000
Activities of daily living (DAD)	372	83 ± 17	82 ± 17	0.450
Medical history				
Smoking, *n* (%)	599	185 (47)	98 (49)	0.640
Hypertension, *n* (%)	616	127 (32)	77 (36)	0.245
Hypercholesterolemia, *n* (%)	616	103 (26)	46 (22)	0.275
Diabetes mellitus, *n* (%)	616	31 (8)	15 (7)	0.770
Cardiovascular disease, *n* (%)	616	71 (18)	53 (25)	0.032
No. of medications	616	2.0 ± 2.0	2.4 ± 2.1	0.062

Abbreviations: APOE Apolipoprotein E, *DAD* Disability Assessment for Dementia (range 0–100)
Years to outcome: in case of alive, follow-up duration; in case of died, duration to death
Data are presented as mean ± SD unless otherwise specified. Group differences were calculated using Student's *t* test for continuous variables. For categorical variables, the chi-square test was used

complaints as a proxy of disease severity (model 3). Effect modification, using interaction terms for each variable with *age and *sex, was not found. Subsequently, we aimed to select the optimal combination of determinants by constructing a multivariate model using forward selection. The model was built by assessing all variables and consecutively selecting the variable with the lowest p value in a stepwise manner until p was < 0.10. In case of several variables with the same lowest p value, we calculated the Wald statistics and selected the variable with the highest Wald value. Variables were added only when the overall model improved, as evaluated using the −2 log-likelihood ratio. In an additional set of analyses, we performed similar analyses based on nonimputed data, and the results were comparable (*see* Additional file 1: Table S1 and S2). Finally, we created Kaplan-Meier curves for each of the variables selected by forward selection. Because all variables except for sex were continuous values, we used tertiles for the survival curves. Data are represented as HRs with accompanying 95% CIs.

Results

Table 1 presents the baseline characteristics of the patients. After a follow-up of 4.9 ± 2.0 years, 213(35%) patients had died (duration baseline AD diagnosis to death 4.3 ± 2.1 years) and 403(65%) patients were alive (follow-up duration 5.3 ± 1.8 years) on the 19th October 2016. Patients who had died were more often male, older and more often had cardiovascular disease. There was no difference in self-reported duration of complaints or years or activities of daily living (as measured with the DAD.

Patients who had died performed worse at baseline on TMT-A and RAVLT immediate recall, but MMSE scores and performance on the other cognitive tests were similar. In addition, these patients' biomarkers were indicative of more severe AD pathology, with a higher MTA and GCA, lower $A\beta_{42}$, and higher p-tau values (Table 2).

We used Cox proportional hazards models to evaluate associations between the individual determinants and mortality, taking into account time to death (Tables 3 and 4). Male sex and older age were associated with an increased risk of mortality. After adjustment for age and sex, worse performance on MMSE, digit span backward, VAT naming, TMT-A, TMT-B, and RAVLT immediate recall and category fluency were associated with an increased risk of mortality. In addition, more severe MTA and GCA seen on MRI scans were associated with an increased risk of mortality. Duration of complaints, activities of daily living (as measured with the DAD), years of education, *APOE* ε4 presence, comorbidity, MRI measures of small vessel disease, and CSF biomarkers were not associated with mortality. When we adjusted additionally for MMSE and duration of complaints as a proxy for disease severity, all related variables from model 2, except MTA, remained associated with mortality.

Table 2 Disease-specific characteristics at baseline, according to outcome

	No. of patients	Alive (n = 403)	Died (n = 213)	p Value
Cognitive tests				
MMSE	616	22 ± 3	22 ± 3	0.480
Digit span forward	597	11 ± 3	11 ± 3	0.908
Digit span backward	593	7 ± 3	6 ± 2	0.154
VAT naming	576	11 ± 1	11 ± 2	0.194
VAT memory	579	6 ± 4	6 ± 4	0.641
TMT-A, seconds	581	81 ± 62	92 ± 64	0.046
TMT-B, seconds	581	299 ± 235	329 ± 215	0.079
RAVLT, immediate recall	551	23 ± 7	22 ± 8	0.026
RAVLT, delayed recall	549	2 ± 2	2 ± 2	0.391
Category fluency	563	13 ± 5	13 ± 6	0.325
MRI				
MTA	484	1.2 ± 0.8	1.6 ± 0.9	0.000
PA	470	1.2 ± 0.8	1.3 ± 0.8	0.185
GCA	482	1.0 ± 0.6	1.2 ± 0.7	0.004
WMH	485	1.0 ± 0.8	1.1 ± 0.9	0.152
Lacunes present, n (%)	483	20 (6)	16 (9)	0.262
Microbleeds by category, n (%)	393			0.064
0 microbleeds		181 (75)	112 (74)	
1–2 microbleeds		41 (17)	17 (11)	
≥ 3 microbleeds		20 (8)	22 (15)	
Infarcts present, n (%)	482	3 (1)	4 (2)	0.257
CSF				
$A\beta_{42}$, pg/ml	466	525 ± 172	490 ± 173	0.037
tau, pg/ml	460	662 ± 340	695 ± 434	0.374
p-tau, pg/ml	463	83 ± 33	91 ± 45	0.031

Abbreviations: MMSE Mini Mental State Examination (score range 0–30), *Digit span forward and backward* (range 0–21), *VAT* Visual Association Test (naming range 0–12, memory range 0–12), *TMT* Trail Making Test (no range), *RAVLT* Rey Auditory Verbal Learning Task (immediate recall range 0–60, delayed recall range 0–15), *MTA* Medial temporal lobe atrophy (range 0–4; average score of left and right sides), *PA* Parietal atrophy (range 0–3; average score of left and right sides), *GCA* Global cortical atrophy (range 0–3), *WMH* White matter hyperintensities (range 0–3), *$A\beta_{42}$* β-Amyloid 1–42, *p-tau* Tau phosphorylated at threonine 181, *MRI* Magnetic resonance imaging, *CSF* Cerebrospinal fluid
Data are presented as mean ± SD unless otherwise specified. Group differences were calculated using Student's t test for continuous variables. For categorical variables, the chi-square test was used

Next, we aimed to identify the optimal combination of determinants in a multivariate model. With use of forward selection, the model included age (HR 1.31, 95% CI 1.12–1.54, p = 0.001), male sex (HR 1.67, 95% CI 1.26–2.21, p = 0.000), digit span backward (HR 1.22,

Table 3 Cox proportional hazards models used to evaluate influence of baseline characteristics and medical history on survival

	Model 1 unadjusted		Model 2 adjusted for age and sex		Model 3: model 2 plus MMSE and duration of complaints	
	HR (95% CI)	*p* value	HR (95% CI)	*p* Value	HR (95% CI)	*p* Value
Demographics						
Male sex[a]	1.57 (1.20–2.07)	0.001	1.60 (1.22–2.11)	0.001	1.79 (1.35–2.37)	0.000
Age	1.27 (1.11–1.46)	0.001	1.29 (1.12–1.48)	0.000	1.33 (1.15–1.53)	0.000
Years of education	0.99 (0.86–1.13)	0.844	0.97 (0.84–1.11)	0.636	1.0 (0.90–11.9)	0.671
Years of complaints	0.88 (0.76–1.03)	0.107	0.88 (0.76–1.03)	0.103	0.87 (0.74–1.01)	0.060
APOE ε4 carrier[a]	0.79 (0.59–1.06)	0.114	0.81 (0.61–1.09)	0.163	0.81 (0.60–1.09)	0.170
Activities of daily living (DAD)[b]	1.13 (0.97–1.31)	0.124	1.11 (0.95–1.29)	0.204	1.09 (0.94–1.26)	0.278
Medical history						
Smoking present[a]	1.18 (0.89–1.55)	0.250	1.09 (0.82–1.45)	0.541	1.12 (0.85–1.49)	0.419
Hypertension present[a]	1.24 (0.94–1.64)	0.130	1.11 (0.83–1.49)	0.467	1.10 (0.82–1.47)	0.528
Hypercholesterolemia present[a]	0.86 (0.62–1.19)	0.861	0.73 (0.52–1.01)	0.059	0.75 (0.54–1.05)	0.091
Diabetes mellitus present[a]	0.72 (0.43–1.22)	0.228	0.62 (0.37–1.06)	0.079	0.65 (0.40–1.05)	0.108
Cardiovascular disease present[a]	1.35 (0.99–1.84)	0.060	1.07 (0.77–1.48)	0.700	1.09 (0.78–1.51)	0.625
No. of medications	1.17 (1.03–1.33)	0.017	1.08 (0.95–1.24)	0.251	1.11 (0.97–1.28)	0.125

Abbreviations: APOE Apolipoprotein E, *DAD* Disability Assessment for Dementia, *MMSE* Mini Mental State Examination
Data are presented as HR (95% CI) using pooled data of 15 imputations per SD increase for continuous variables or for the presence of the dichotomous variable for mortality
[a] Dichotomous variable
[b] Because a lower score indicates worse performance, these scores were inverted

95% CI 1.03–1.43, p = 0.018), TMT-A (HR 1.22, 95% CI 1.06–1.41, p = 0.005), MTA (HR 1.18, 95% CI 1.01–1.38, p = 0.038), and CSF p-tau (HR 1.15, 95% CI 1.00–1.32, p = 0.058). Survival curves for these variables are shown in Fig. 1. Of note, because < 50% of our subjects had died, median survival time can only be estimated from these curves.

Discussion

Our main finding is that despite their relatively young age, roughly one of three patients with AD had died with a mean of 5 years after diagnosis. Predisposing factors for an increased risk of mortality were older age, male sex, more severe executive dysfunction, presence of MTA, and higher p-tau in CSF, indicative of more severe AD pathology. By contrast, duration of complaints, level of activities of daily living, *APOE* ε4 status, and comorbidity were not related to mortality.

In our relatively young population (average age 67 ± 8 years) derived from a tertiary memory clinic cohort with mild to moderate dementia (all with MMSE scores > 16; average MMSE score 22), 35% of the patients had died within 5 years after receiving their baseline diagnosis. This mortality rate is considerably higher than that of the general Dutch population [19]. Previous studies described slightly higher mortality rates, but most studies included patients older than 75 years of age or with severe dementia with MMSE scores < 20 [38–40]. Only a few studies have been focused on mortality in young patients with AD or less affected patients, showing comparable mortality rates [38, 41–43]. Extending these former studies, we evaluated not only comorbidity but also focused on disease-specific markers as determinants of mortality.

In addition to male sex and older age, both of which are known determinants of mortality in AD and in the general population, we found executive dysfunction, MTA, and higher p-tau in CSF, all reflecting more severe disease, to be determinants of mortality. Self-reported duration of complaints was not associated with mortality, indicating that the patients who died had more aggressive rather than more advanced disease. In line with this notion, there was hardly any difference at baseline in the severity of cognitive impairment between those who died within the study period and those who remained alive. Previous studies focused on neuropsychology have shown mainly an association with mortality when assessing decline over time but not at baseline [5, 38, 39]. In our study, we consistently found tests in the executive domain and, to a lesser degree, memory as determinants of mortality. A potential explanation for this finding is that subjects with executive dysfunction are at greater risk of dependency, increasing the risk of complications. Also, the executive domain seems to be a mediator for other cognitive domains, whereas tests for delayed recall were already at floor level in many patients [44]. This latter finding could explain why tests for delayed recall showed no association with survival.

Table 4 Cox proportional hazards models used to evaluate influence of cognitive performance, magnetic resonance imaging, and cerebrospinal fluid at baseline on survival

	Model 1 unadjusted		Model 2 adjusted for age and sex		Model 3: model 2 plus MMSE and duration of complaints	
	HR (95% CI)	*p* Value	HR (95% CI)	*p* Value	HR (95% CI)	*p* Value
Cognitive tests						
MMSE[a]	1.11 (0.97–1.28)	0.131	1.23 (1.07–1.42)	0.005	1.25 (1.08–1.44)	0.003
Digit span forward[a]	1.07 (0.93–1.23)	0.362	1.10 (0.95–1.26)	0.207	1.03 (0.89–1.20)	0.651
Digit span backward[a]	1.21 (1.05–1.40)	0.008	1.31 (1.13–1.52)	0.000	1.24 (1.06–1.46)	0.009
VAT naming[a]	1.15 (1.01–1.31)	0.037	1.14 (1.01–1.30)	0.042	1.11 (0.97–1.27)	0.136
VAT memory[a]	1.01 (0.88–1.15)	0.937	1.07 (0.93–1.23)	0.360	1.02 (0.88–1.19)	0.790
TMT-A	1.21 (1.07–1.37)	0.003	1.29 (1.14–1.47)	0.000	1.23 (1.08–1.41)	0.003
TMT-B	1.19 (1.05–1.35)	0.006	1.28 (1.13–1.45)	0.000	1.21 (1.06–1.40)	0.005
RAVLT, immediate recall[a]	1.23 (1.06–1.43)	0.008	1.19 (1.02–1.38)	0.025	1.11 (0.95–1.30)	0.193
RAVLT, delayed recall[a]	0.96 (0.84–1.10)	0.533	0.96 (0.83–1.10)	0.507	0.90 (0.78–1.04)	0.154
Category fluency[a]	1.17 (1.02–1.37)	0.045	1.17 (1.01–1.36)	0.041	1.10 (0.94–1.29)	0.243
MRI						
MTA	1.26 (1.10–1.44)	0.001	1.18 (1.02–1.37)	0.030	1.15 (0.98–1.34)	0.081
PA	1.12 (0.97–1.30)	0.113	1.10 (0.95–1.28)	0.192	1.12 (0.96–1.29)	0.143
GCA	1.21 (1.05–1.40)	0.008	1.18 (1.01–1.37)	0.037	1.17 (1.00–1.36)	0.044
WMH	1.16 (1.01–1.33)	0.041	1.07 (0.92–1.25)	0.364	1.05 (0.90–1.22)	0.518
Lacunes present[b]	1.15 (0.77–1.71)	0.505	1.10 (0.73–1.66)	0.634	1.17 (0.76–1.79)	0.485
Microbleed categories						
Microbleeds, 1–2	0.82 (0.49–1.37)	0.450	0.72 (0.43–1.19)	0.195	0.69 (0.42–1.16)	0.163
Microbleeds, ≥ 3	1.09 (0.80–1.47)	0.598	1.03 (0.76–1.40)	0.840	1.01 (0.74–1.37)	0.956
Infarcts present[b]	1.11 (0.63–2.00)	0.710	1.15 (0.64–2.05)	0.641	1.11 (0.60–2.05)	0.727
CSF						
$A\beta_{42}^{a}$	0.98 (0.84–1.14)	0.765	1.02 (0.87–1.18)	0.850	0.99 (0.86–1.16)	0.943
tau	1.05 (0.91–1.22)	0.504	1.09 (0.94–1.27)	0.275	1.07 (0.92–1.26)	0.369
p-tau	1.06 (0.92–1.23)	0.426	1.09 (0.94–1.26)	0.242	1.08 (0.93–1.26)	0.316

Abbreviations: MMSE Mini Mental State Examination, *VAT* Visual Association Test, *TMT* Trail Making Test, *RAVLT* Rey Auditory Verbal Learning Task, *MRI* Magnetic resonance imaging, *MTA* Medial temporal lobe atrophy, *PA* Parietal atrophy, *GCA* Global cortical atrophy, *WMH* White matter hyperintensities score 3, *CSF* Cerebrospinal fluid, $A\beta_{42}$ β-Amyloid 1–42, *p-tau* Tau phosphorylated at threonine 181

Data are presented as HR (95% CI) using pooled data of 15 imputations per SD increase for continuous variables or for the presence of the dichotomous variable for mortality

[a] Because a lower score indicates worse performance, these scores were inverted

[b] Dichotomous variable

To our knowledge in only two other studies have researchers assessed associations of MRI atrophy markers with mortality, with findings that global atrophy, but not MTA, was associated with mortality in dementia [12, 18]. An association of MTA and mortality in AD was found in a study conducted with computed tomographic scans [45]. In our univariate models, we found more severe MTA and global atrophy associated with increased risk of death; in the multivariate forward selection model, GCA was not included. Atrophy is seen on MRI scans as a marker of downstream neuronal degeneration [25]. In this study, other markers of neurodegeneration, such as p-tau in CSF, were also included in our multivariate forward selection model, which confirms the results of the few studies addressing CSF and mortality in AD [16, 18]. The effect of WMH seems attributable largely to age, because the prognostic value disappeared in the adjusted models. This is different from what has been found before and could potentially be explained by the relatively young age of our sample [12].

In line with previous studies, male sex was associated with higher mortality in AD [3]. It has been suggested that women present earlier in their disease course owing to more easily noticed impairment in household tasks, and hence they have a longer survival time [3]. Also, women more often lived alone and were more frequently widowed, leading to impairment being noticed earlier. We did not find an association with level of activities of

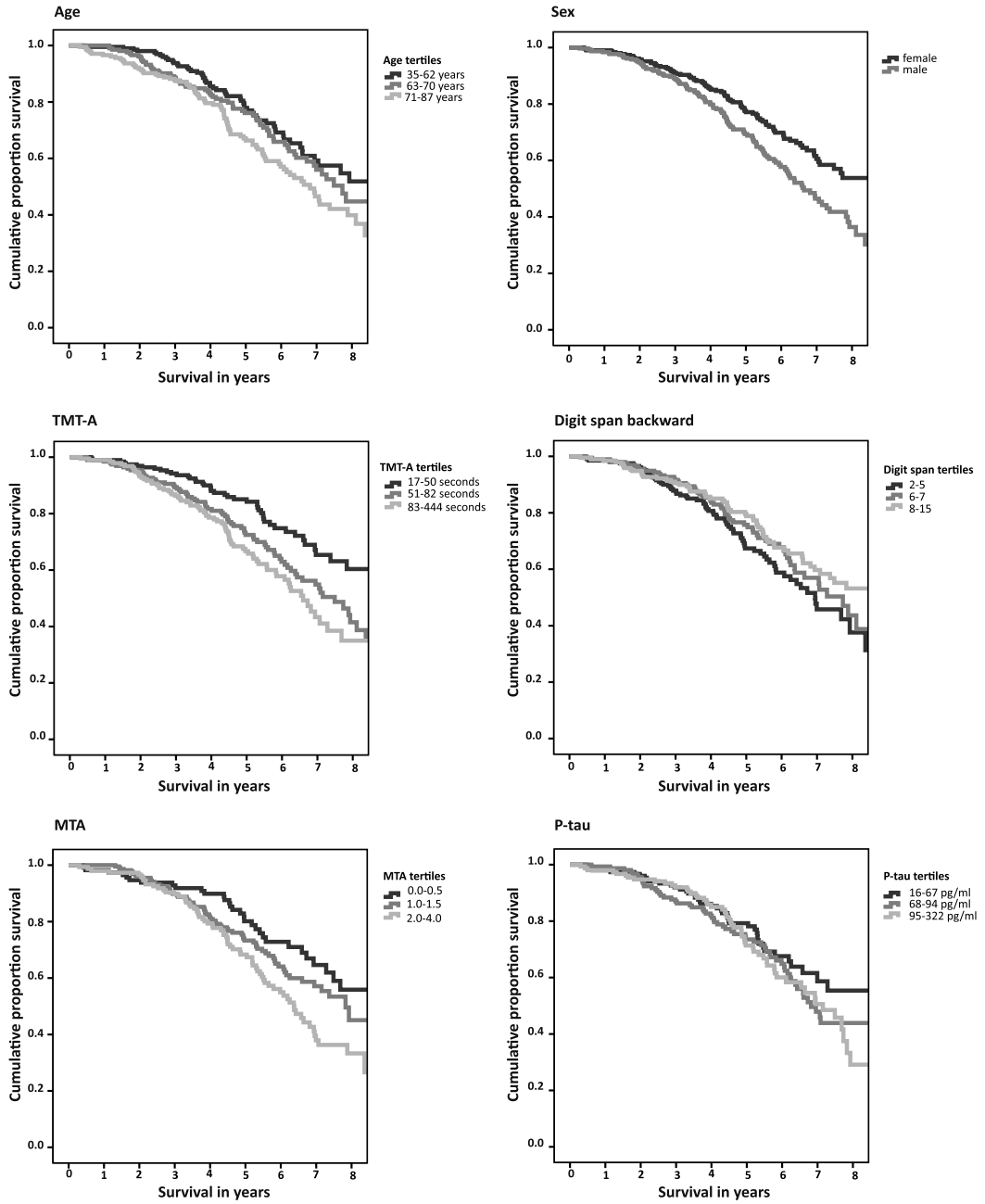


Fig. 1 Kaplan-Meier curve, according to variables from forward selection model: age, sex, TMT-A, digit span backward, MTA, and p-tau (all except sex stratified in tertiles). Legend: Note: digit span backward: range 0-21, TMT: trail making test (no range), MTA: medial temporal lobe atrophy ranging 0-4 (average score of left and right side), p-tau: tau phosphorylated at threonine 181. Survival curves were calculated using raw data, without imputation

daily living (as measured by the DAD). We believe this is possibly most relevant in more advanced disease stages and not in our cohort, where activities of daily living were only mildly impaired in most patients [41, 46]. Finally, and contrary to our expectations, we could not confirm smoking, comorbidity, or number of medications as predisposing factors for an increased risk of mortality. Previous studies have shown an association of cardiovascular risk factors with mortality, but these studies were focused mostly on older populations that are by definition at higher risk of both cardiovascular disease and mortality [11, 13–15, 46]. Also, a higher level of comorbidity has previously been shown to relate to survival time [14, 15], but again in older populations; in our present study, we used number of medications as a proxy of level of comorbidity and found no association [13]. Our study shows that the AD process itself, as reflected by neuropsychology as well as MRI and CSF biomarkers, has prognostic value in terms of mortality as well. This fits with the observation that patients with AD have higher rates of mortality than the general population and that AD is the swiftest growing cause of death in the Western world [2, 19].

Limitations of the present study are that our population was derived from a tertiary memory clinic, which hampers the generalizability of the results. However, the added value of our study is its focus on younger patients, for whom a paucity of data exists. We studied a broad range of determinants in patients with a relatively long follow-up duration. In addition, we included only patients with MMSE scores ≥ 16 to prevent cognitive testing from being at floor level. Of note, even in our young, mild to moderately impaired cohort of patients with AD, mortality was high. Another limitation might be the mean follow-up duration of 5.3 ± 1.8 years for the patients who remained alive, implying that these patients might have died shortly after this period. Nonetheless, all patients had a minimum follow-up of 2 years. Finally, we had information on medication use only at baseline and thus had no information on the prescription of cholinesterase inhibitors after the diagnosis AD. This could be a limitation because some studies have shown that cholinesterase inhibitors can increase survival, whereas others have shown no such effect or only in older patients [41, 43, 47]. Furthermore, we were not able to look at the relationship between use of antipsychotics and mortality [48, 49], because only a very small proportion of our subjects used these medications. However, use of antipsychotics is likely to occur later in the course of the disease. Among the strengths of the present study is our harmonized diagnostic protocol according to which all patients were assessed, because all patients were selected from the same memory clinic and received the same diagnostic workup and similar treatment and management.

Conclusions

Our results have important clinical implications. We found that AD-related factors, rather than comorbidity or duration of complaints, were associated with increased mortality in our relatively young cohort. This knowledge enables timely dialogue on prognosis, even in young patients who are otherwise healthy.

Additional file

Additional file 1: Table S1. Cox proportional hazard models using nonimputed data; influence of baseline characteristics and medical history on survival status. **Table S2.** Cox proportional hazard models based on nonimputed data; influence of CSF and MRI on survival status. (DOCX 26 kb)

Abbreviations

$A\beta_{42}$: β-Amyloid 1–42; AD: Alzheimer's disease; *APOE*: Apolipoprotein E; CSF: Cerebrospinal fluid; DAD: Disability Assessment for Dementia; FLAIR: Fluid-attenuated inversion recovery; GCA: Global cortical atrophy; MMSE: Mini Mental State Examination; MRI: Magnetic resonance imaging; MTA: Medial temporal lobe atrophy; PA: Parietal atrophy; p-tau: Tau phosphorylated at threonine 181; RAVLT: Rey Auditory Verbal Learning Task; TMT-A: Trail Making Test A; TMT-B: Trail Making Test B; VAT: Visual Association Test; WMH: White matter hyperintensities

Acknowledgements

Research done at the VUmc Alzheimer Center is part of the neurodegeneration research program of Amsterdam Neuroscience. The VUmc Alzheimer Center is supported by Alzheimer Nederland and Stichting VUmc Fonds. The clinical database structure was developed with funding from Stichting Dioraphte. HFMRM is appointed on a grant from the European Union Seventh Framework Program project PredictND under grant agreement 611005.

Funding

Not applicable.

Availability of data and materials

The datasets used and/or analyzed during the present study are available from the corresponding author on reasonable request.

Authors' contributions

HFMRM drafted the manuscript and analyzed/interpreted data. HL revised the manuscript and analyzed/interpreted data. TK revised the manuscript and interpreted data. AWL revised the manuscript and interpreted data. CET revised the manuscript and interpreted data. FB revised the manuscript and interpreted data. PS revised the manuscript and interpreted data. MvG revised the manuscript and interpreted data. JL revised the manuscript and analyzed/interpreted data. WMvdF drafted the manuscript, analyzed/interpreted data, and supervised the project. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The study was approved by the local medical ethics committee. All patients provided written informed consent for their clinical data to be used for research purposes.

Consent for publication

Not applicable.

Competing interests

CET is a member of the Innogenetics International Advisory Boards of Fujirebio/Innogenetics and Roche. FB serves/has served on the advisory boards of Bayer-Schering Pharma, Sanofi-Aventis, Biogen-Idec, Teva Pharmaceutical Industries, Merck-Serono, Novartis, Roche, Synthon BV, Jansen Research, and Genzyme. FB has received funding from the Dutch MS Society and the European Union Seventh Framework Program (EU-FP7) and has been a speaker at symposia

organized by the Serono Symposia Foundation and Medscape. PS has served as a consultant for Wyeth-Elan, Genentech, Danone, and Novartis and has received funding for travel from Pfizer, Elan, Janssen, and Danone Research. JL reports that Combinostics Oy owns the following intellectual property rights related to this paper: (1) J. Koikkalainen and J. Lotjonen. Method for inferring the state of a system. Publication number US 7840510 B2; application number PCT/FI2007/050277; (2) J. Lotjonen, J. Koikkalainen, and J. Mattila. State inference in a heterogeneous system. Application number PCT/FI2010/050545; FI20125177. JL is a shareholder in Combinostics Oy. WMvdF performs contract research for Boehringer Ingelheim. Research programs of WMvdF have been funded by ZonMw, the Netherlands Organization for Scientific Research (now), EU-FP7, Alzheimer Nederland, Cardiovascular Onderzoek Nederland, Stichting Dioraphte, Gieskes-Strijbis Fonds, Boehringer Ingelheim, Piramal Neuroimaging, Roche BV, and Janssen Stellar. All funding is paid to WMvdF's institution. The other authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Alzheimer Center, Department of Neurology, VU University Medical Center, Amsterdam Neuroscience, P.O. Box 7057, 1007, MB, Amsterdam, The Netherlands. [2]VTT Technical Research Center of Finland Ltd, Tampere, Finland. [3]Department of Medical Psychology, VU University Medical Center, Amsterdam Neuroscience, Amsterdam, The Netherlands. [4]Neurochemistry Lab and Biobank, Department of Clinical Chemistry, VU University Medical Center, Amsterdam Neuroscience, Amsterdam, The Netherlands. [5]Department of Radiology and Nuclear Medicine, VU University Medical Center, Amsterdam Neuroscience, Amsterdam, The Netherlands. [6]Institute of Neurology, UCL, London, UK. [7]Institute of Healthcare Engineering, UCL, London, UK. [8]Combinostics Ltd, Tampere, Finland. [9]Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam Neuroscience, Amsterdam, The Netherlands.

Received: 9 November 2017 Accepted: 22 January 2018
Published online: 20 February 2018

References

1. World Health Organization. Dementia fact sheet (No. 362). http://www.who.int/mediacentre/factsheets/fs362/en/. Accessed 11 Oct 2016.
2. Brodaty H, Seeher K, Gibson L. Dementia time to death: a systematic literature review on survival time and years of life lost in people with dementia. Int Psychogeriatr. 2012;24:1034–45.
3. Todd S, Barr S, Roberts M, Passmore AP. Survival in dementia and predictors of mortality: a review. Int J Geriatr Psychiatry. 2013;28:1109–24.
4. Wattmo C, Londos E, Minthon L. Risk factors that affect life expectancy in Alzheimer's disease: a 15-year follow-up. Dement Geriatr Cogn Disord. 2014;38:286–99.
5. Leoutsakos JM, Forrester SN, Corcoran CD, Norton MC, Rabins PV, Steinberg MI, Tschanz JT, Lyketsos CG. Latent classes of course in Alzheimer's disease and predictors: the Cache County Dementia Progression Study. Int J Geriatr Psychiatry. 2015;30:824–32.
6. Rabins PV, Schwartz S, Black BS, Corcoran C, Fauth E, Mielke M, Christensen J, Lyketsos C, Tschanz J. Predictors of progression to severe Alzheimer's disease in an incidence sample. Alzheimers Dement. 2013;9:204–7.
7. Spalletta G, Long JD, Robinson RG, Trequattrini A, Pizzoli S, Caltagirone C, Orfei MD. Longitudinal neuropsychiatric predictors of death in Alzheimer's disease. J Alzheimers Dis. 2015;48:627–36.
8. Tschanz JT, Corcoran CD, Schwartz S, Treiber K, Green RC, Norton MC, Mielke MM, Piercy K, Steinberg M, Rabins PV, et al. Progression of cognitive, functional, and neuropsychiatric symptom domains in a population cohort with Alzheimer dementia: the Cache County Dementia Progression Study. Am J Geriatr Psychiatry. 2011;19:532–42.
9. van de Vorst IE, Koek HL, de Vries R, Bots ML, Reitsma JB, Vaartjes I. Effect of vascular risk factors and diseases on mortality in individuals with dementia: a systematic review and meta-analysis. J Am Geriatr Soc. 2016;64:37–46.
10. Koedam EL, Pijnenburg YA, Deeg DJ, Baak MM, van der Vlies AE, Scheltens P, van der Flier WM. Early-onset dementia is associated with higher mortality. Dement Geriatr Cogn Disord. 2008;26:147–52.
11. van de Vorst IE, Vaartjes I, Geerlings MI, Bots ML, Prognosis KHL. of patients with dementia: results from a prospective nationwide registry linkage study in the Netherlands. BMJ Open. 2015;5:e008897.
12. Henneman WJ, Sluimer JD, Cordonnier C, Baak MM, Scheltens P, Barkhof F, van der Flier WM. MRI biomarkers of vascular damage and atrophy predicting mortality in a memory clinic population. Stroke. 2009;40:492–8.
13. Garcia-Ptacek S, Farahmand B, Kareholt I, Religa D, Cuadrado ML, Eriksdotter M. Mortality risk after dementia diagnosis by dementia type and underlying factors: a cohort of 15,209 patients based on the Swedish Dementia Registry. J Alzheimers Dis. 2014;41:467–77.
14. Larson EB, Shadlen MF, Wang L, McCormick WC, Bowen JD, Teri L, Kukull WA. Survival after initial diagnosis of Alzheimer disease. Ann Intern Med. 2004;140:501–9.
15. Zekry D, Herrmann FR, Graf CE, Giannelli S, Michel JP, Gold G, High KKH. levels of comorbidity and disability cancel out the dementia effect in predictions of long-term mortality after discharge in the very old. Dement Geriatr Cogn Disord. 2011;32:103–10.
16. Degerman Gunnarsson M, Lannfelt L, Ingelsson M, Basun H, Kilander L. High tau levels in cerebrospinal fluid predict rapid decline and increased dementia mortality in Alzheimer's disease. Dement Geriatr Cogn Disord. 2014;37:196–206.
17. Wallin AK, Blennow K, Zetterberg H, Londos E, Minthon L, Hansson O. CSF biomarkers predict a more malignant outcome in Alzheimer disease. Neurology. 2010;74:1531–7.
18. Nagga K, Wattmo C, Zhang Y, Wahlund LO, Palmqvist S. Cerebral inflammation is an underlying mechanism of early death in Alzheimer's disease: a 13-year cause-specific multivariate mortality study. Alzheimers Res Ther. 2014;6:41.
19. Centraal Bureau voor de Statistiek. Statline [updated 2003]. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/83795NED/table?ts=1517513589647. Accessed 3 Jan 2018.
20. van der Vlies AE, Koedam EL, Pijnenburg YA, Twisk JW, Scheltens P, van der Flier WM. Most rapid cognitive decline in APOE ε4 negative Alzheimer's disease with early onset. Psychol Med. 2009;39:1907–11.
21. Ueki A, Shinjo H, Shimode H, Nakajima T, Morita Y. Factors associated with mortality in patients with early-onset Alzheimer's disease: a five-year longitudinal study. Int J Geriatr Psychiatry. 2001;16:810–5.
22. Palmqvist S, Hertze J, Minthon L, Wattmo C, Zetterberg H, Blennow K, Londos E, Hansson O. Comparison of brief cognitive tests and CSF biomarkers in predicting Alzheimer's disease in mild cognitive impairment: six-year follow-up study. PLoS One. 2012;7:e38639.
23. van der Flier WM, Pijnenburg YA, Prins N, Lemstra AW, Bouwman FH, Teunissen CE, van Berckel BN, Stam CJ, Barkhof F, Visser PJ, et al. Optimizing patient care and research: the Amsterdam Dementia Cohort. J Alzheimers Dis. 2014;41:313–27.
24. Gelinas I, Gauthier L, McIntyre M, Gauthier S. Development of a functional measure for persons with Alzheimer's disease: the disability assessment for dementia. Am J Occup Ther. 1999;53:471–81.
25. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack Jr CR, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 2011;7:263–9.
26. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology. 1984;34:939–44.
27. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975;12:189–98.
28. Lindeboom J, Schmand B, Tulner L, Walstra G, Jonker C. Visual Association. Test to detect early dementia of the Alzheimer type. J Neurol Neurosurg Psychiatry. 2002;73:126–33.
29. Schmand B, Houx P, de Koning I. Norms of psychological tests for usage in the clinical neuropsychology [in Dutch]. Utrecht, the Netherlands: Dutch Association of Psychologists; 2012. https://www.psynip.nl/sectorensecties/sector-gezondheidszorg/neuropsychologie/. Accessed 11 Oct 2016.
30. Reitan R. Validity of the Trail Making Test as an indicator of organic brain damage. Percept Mot Skills. 1958;8:271–6.
31. Lindeboom J, Matto D. Digit series and Knox cubes as concentration tests for elderly subjects [in Dutch]. Tijdschr Gerontol Geriatr. 1994;25:63–8.

32. Van der Elst W, Van Boxtel MP, Van Breukelen GJ, Jolles J. Normative data for the Animal, Profession and Letter M Naming verbal fluency tests for Dutch speaking participants and the effects of age, education, and sex. J Int Neuropsychol Soc. 2006;12:80–9.
33. Scheltens P, Launer LJ, Barkhof F, Weinstein HC, van Gool WA. Visual assessment of medial temporal lobe atrophy on magnetic resonance imaging: interobserver reliability. J Neurol. 1995;242:557–60.
34. Pasquier F, Leys D, Weerts JG, Mounier-Vehier F, Barkhof F, Scheltens P. Inter- and intraobserver reproducibility of cerebral atrophy assessment on MRI scans with hemispheric infarcts. Eur Neurol. 1996;36:268–72.
35. Koedam EL, Lehmann M, van der Flier WM, Scheltens P, Pijnenburg YA, Fox N, Barkhof F, Wattjes MP. Visual assessment of posterior atrophy development of a MRI rating scale. Eur Radiol. 2011;21:2618–25.
36. Fazekas F, Chawluk JB, Alavi A, Hurtig HI, Zimmerman RA. MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. AJR Am J Roentgenol. 1987;149:351–6.
37. Duits FH, Teunissen CE, Bouwman FH, Visser PJ, Mattsson N, Zetterberg H, Blennow K, Hansson O, Minthon L, Andreasen N, et al. The cerebrospinal fluid "Alzheimer profile": easily said, but what does it mean? Alzheimers Dement. 2014;10:713–23. e2
38. Hui JS, Wilson RS, Bennett DA, Bienias JL, Gilley DW, Evans DA. Rate of cognitive decline and mortality in Alzheimer's disease. Neurology. 2003;61:1356–61.
39. Wilson RS, Li Y, Aggarwal NT, McCann JJ, Gilley DW, Bienias JL, Barnes LL, Evans DA. Cognitive decline and survival in Alzheimer's disease. Int J Geriatr Psychiatry. 2006;21:356–62.
40. Soto ME, Andrieu S, Cantet C, Reynish E, Ousset PJ, Arbus C, Gillette-Guyonnet S, Nourhashemi F, Vellas B. Predictive value of rapid decline in Mini Mental State Examination in clinical practice for prognosis in Alzheimer's disease. Dement Geriatr Cogn Disord. 2008;26:109–16.
41. Wattmo C, Londos E, Minthon L. Longitudinal associations between survival in Alzheimer's disease and cholinesterase inhibitor use, progression, and community-based services. Dement Geriatr Cogn Disord. 2015;40:297–310.
42. Morris JC, Edland S, Clark C, Galasko D, Koss E, Mohs R, van Belle G, Fillenbaum G, Heyman A. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD): Part IV. Rates of cognitive change in the longitudinal assessment of probable Alzheimer's disease. Neurology. 1993;43:2457–65.
43. Gillette-Guyonnet S, Andrieu S, Nourhashemi F, Gardette V, Coley N, Cantet C, Gauthier S, Ousset PJ, Vellas B. Long-term progression of Alzheimer's disease in patients under antidementia drugs. Alzheimers Dement. 2011;7:579–92.
44. Pillai JA, Bonner-Jackson A, Walker E, Mourany L, Cummings JL. Higher working memory predicts slower functional decline in autopsy-confirmed Alzheimer's disease. Dement Geriatr Cogn Disord. 2014;38:224–33.
45. Claus JJ, van Gool WA, Teunisse S, Walstra GJ, Kwa VI, Hijdra A, Verbeeten Jr B, Koelman JH, Bour LJ, Ongerboer De Visser BW. Predicting survival in patients with early Alzheimer's disease. Dement Geriatr Cogn Disord. 1998;9:284–93.
46. Rountree SD, Chan W, Pavlik VN, Darby EJ, Doody RS. Factors that influence survival in a probable Alzheimer disease cohort. Alzheimers Res Ther. 2012;4:16.
47. Nordstrom P, Religa D, Wimo A, Winblad B, Eriksdotter M. The use of cholinesterase inhibitors and the risk of myocardial infarction and death: a nationwide cohort study in subjects with Alzheimer's disease. Eur Heart J. 2013;34:2585–91.
48. Scarmeas N, Brandt J, Albert M, Hadjigeorgiou G, Papadimitriou A, Dubois B, Sarazin M, Devanand D, Honig L, Marder K, et al. Delusions and hallucinations are associated with worse outcome in Alzheimer disease. Arch Neurol. 2005;62:1601–8.
49. Nielsen RE, Lolk A, Valentin JB, Andersen K. Cumulative dosages of antipsychotic drugs are associated with increased mortality rate in patients with Alzheimer's dementia. Acta Psychiatr Scand. 2016;134:314–20.

Additional file 1

Table S1 Cox proportional hazard models, using non-imputed data; influence of baseline characteristics and medical history on survival status.

		Model 1 *unadjusted*		**Model 2** *adjusted for age and sex*		**Model 3** *model 2 plus MMSE and duration of complaints*	
		HR (95% CI)	***p-value***	**HR (95% CI)**	***p-value***	**HR (95% CI)**	***p-value***
Demographics	**Gender, male**[a]	1.57(1.20-2.10))	.001	1.61(1.22-2.11)	.001	1.76(1.33-2.33)	.000
	Age	1.27(1.11-1.46)	.001	1.29(1.11-1.48)	.000	1.34(1.16-1.55)	.000
	Years of education	0.99(0.86-1.1.3)	.844	0.97(0.84-1.11)	.636	1.03(0.90-1.19)	.671
	Years of complaints	0.88(0.75-1.20)	.088	0.88(0.76-1.02)	.089	0.86(0.74-1.00)	.050
	APOE e4 carrier [a]	0.79(0.59-1.06)	.109	0.81(0.60-1.09)	.168	0.81(0.60-109)	.166
	Activities of daily living (DAD) [b]	1.05(0.86-1.29)	.623	1.04(0.85-1.28)	.690	1.06(0.86-1.31)	.577
Medical history of	**Smoking present**[a]	1.18(0.89-1.55)	.249	1.10(0.83-1.46)	.529	1.12(0.84-1.49)	.440
	Hypertension present[a]	1.24(0.94-1.65)	.130	1.11(0.83-1.49)	.467	1.08(0.81-1.45)	.608
	Hypercholesterolemia present[a]	0.86(0.62-1.19)	.369	0.73(0.52-1.01)	.059	0.74(0.53-1.03)	.074
	Diabetes mellitus present[a]	0.72(0.43-1.22)	.228	0.62(0.37-1.06)	.079	0.64(0.37-1.08)	.095
	Cardiovascular disease present[a]	1.35(0.99-1.84)	.060	1.07(0.77-1.48)	.700	1.07(0.77-1.49)	.686
	Number of medications	1.17(1.03-1.33)	.017	1.08(0.95-1.24)	.251	1.10(0.96-1.27)	.167

Note: DAD: disability assessment of dementia. Data are presented as hazard ratio (HR) (95% CI) using non-imputed data, per standard deviation increase for continuous variables or for the presence of the dichotomous variable ([a]). [b] since a lower score indicates a worse performance, this score was inverted

Table S2 Cox proportional hazard models based on non-imputed data; influence of CSF and MRI on survival status.

		Model 1 *unadjusted*		**Model 2** *adjusted for age and sex*		**Model 3** *model 2 plus MMSE and duration of complaints*	
		HR (95% CI)	***p-value***	**HR (95% CI)**	***p-value***	**HR (95% CI)**	***p-value***
Neuropsychology	**MMSE[a]**	1.11(0.97-1.28)	.131	1.23(1.07-1.42)	.005	1.26(1.09-1.45)	.002
	Digit span forward[a]	1.07(0.92-1.22)	.380	1.09(0.95-1.26)	.231	1.03(0.89-1.20)	.672
	Digit span backward[a]	1.22(1.05-1.41)	.008	1.31(1.12-1.52)	.001	1.24(1.05-1.46)	.010
	VAT naming[a]	1.15(1.01-1.30)	.034	1.14(1.00-1.29)	.045	1.11(0.97-1.26)	.122
	VAT memory[a]	1.01(0.87-1.16)	.915	1.07(0.92-1.24)	.370	1.04(0.88-1.22)	.647
	TMT-A	1.21(1.06-1.37)	.004	1.29(1.13-1.47)	.000	1.25(1.09-1.433)	.002
	TMT-B	1.18(1.03-1.34)	.014	1.26(1.10-1.44)	.001	1.22(1.06-1.40)	.006
	RAVLT, immediate recall[a]	1.20(1.02-1.41)	.025	1.16(0.99-1.36)	.072	1.08(0.91-1.28)	.385
	RAVLT, delayed recall[a]	0.96(0.83-1.12)	.603	0.96(0.83-1.11)	.574	0.90(0.77-1.06)	.203
	Category fluency[a]	1.14(0.97-1.33)	.113	1.14(0.98-1.34)	.092	1.09(0.93-1.28)	.299
MRI	**MTA**	1.28(1.11-1.47)	.001	1.21 (1.03-1.40)	.017	1.17(1.00-1.37)	.054
	PA	1.18(1.01-1.37)	.039	1.16 (0.99-1.35)	.066	1.17(0.99-1.37)	.056
	GCA	1.23(1.06-1.43)	.006	1.20(1.03-1.40)	.022	1.19(1.02-1.39)	.032
	WMH	1.17(1.01-1.34)	.033	1.08(0.93-1.26)	.322	1.07(0.92-1.25)	.387

	Lacunes present[b]	1.34(0.80-2.25)	.260	1.21(0.72-2.03)	.479	1.33(0.79-2.25)	.283
	Microbleeds, categories						
	Microbleeds, 1-2	0.82(0.49-1.36)	.439	0.70(0.42-1.17)	.172	0.68(0.41-1.15)	.150
	Microbleeds, ≥3	1.71(1.08-2.70)	.022	1.53(0.96-2.44)	.072	1.49(0.94-2.38)	.091
	Infarcts present[b]	1.72(0.64-4.65)	.284	1.89(0.70-5.12)	.210	1.84(0.68-4.99)	.232
CSF	**AB42**[a]	1.01 (0.86-1.20)	.866	1.03(0.88-1.21)	.697	1.01(0.86-1.18)	.897
	t-tau	1.13(0.99-1.29)	.074	1.17 (1.03-1.34)	.017	1.17(1.02-1.34)	.025
	p-tau	1.14(0.99-1.31)	.077	1.16(1.01-1.33)	.035	1.15(1.00-1.32)	.049

Note: MMSE: mini mental state examination VAT: visual association test, TMT: trail making test, RAVLT: Rey auditory verbal learning task, MTA: medial temporal lobe atrophy, PA: patrophy, GCA: global cortical atrophy, WMH: white matter hyperintensities 3, AB42: beta amyloid 1-42, p-tau: tau phosphorylated at threonine 181.

[a]: since a lower score indicates a worse performance, these scores were inverted. Data are presented as hazard ratio (HR) (95% CI) using non-imputed data, per standard deviation increase for continuous variables or for the presence of the dichotomous variable ([b]), for mortality.