Computerized Medical Imaging and Graphics 90 (2021) 101910

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

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

A predictive framework based on brain volume trajectories enabling early detection of Alzheimer's disease

Samaneh Abolpour Mofrad [a,c,*], Arvid Lundervold [b,c], Alexander Selvikvåg Lundervold [a,c], for the Alzheimer's Disease Neuroimaging InitiativeData used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf, for the Australian Imaging Biomarkers and Lifestyle Flagship Study of AgeingData 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. The AIBL researchers contributed data but did not participate in analysis or writing of this report. AIBL researchers are listed at www.aibl.csiro.au.

[a] Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, Postbox 7030, 5020 Bergen, Norway
[b] The Neural Networks and Microcircuits Research Group, Department of Biomedicine, University of Bergen, Bergen, Norway
[c] The Mohn Medical Imaging and Visualization Centre (MMIV), Department of Radiology, Haukeland University Hospital, Bergen, Norway

ARTICLE INFO

Keywords:
Mixed effects models
Machine learning
Longitudinal data analysis
Alzheimer's disease
Mild cognitive impairment
MRI

ABSTRACT

We present a framework for constructing predictive models of cognitive decline from longitudinal MRI examinations, based on mixed effects models and machine learning. We apply the framework to detect conversion from cognitively normal (CN) to mild cognitive impairment (MCI) and from MCI to Alzheimer's disease (AD), using a large collection of subjects sourced from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the Australian Imaging, Biomarkers and Lifestyle Flagship Study of Aging (AIBL). We extract subcortical segmentation and cortical parcellation from corresponding T1-weighted images using FreeSurfer v.6.0, select bilateral 3D regions of interest relevant to neurodegeneration/dementia, and fit their longitudinal volume trajectories using linear mixed effects models. Features describing these model-based trajectories are then used to train an ensemble of machine learning classifiers to distinguish stable CN from converters to MCI, and stable MCI from converters to AD. On separate test sets the models achieved an average of accuracy/precision/recall score of 69/73/60% for converted to MCI and 75/74/77% for converted to AD, illustrating the framework's ability to extract predictive imaging-based biomarkers from routine T1-weighted MRI acquisitions.

1. Introduction

About 50 million people world-wide suffer from dementia (World Health Organization, 2019), with a new case appearing every 3.2 seconds (Prince, 2015). The total cost of dementia care has risen to above one trillion US dollars after 2018 (World Health Organization, 2019; Prince, 2015). The most common form of dementia is Alzheimer's disease (AD), responsible for up to 60 − 70% of cases (Prince, 2015). AD

* Corresponding author at: Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, Postbox 7030, 5020 Bergen, Norway.
E-mail address: sam@hvl.no (S.A. Mofrad).

https://doi.org/10.1016/j.compmedimag.2021.101910
Received 17 September 2020; Received in revised form 12 February 2021; Accepted 26 March 2021
Available online 2 April 2021

is aging-related, mostly inflicting people above 60 years. This is a steadily growing age group: in the 20th and 21st centuries, both the overall population levels and life expectancy increased drastically, a trend that shows no sign of stopping. In 2018, for the first time recorded in history, people aged 65 and older outnumbered children five years or younger. Currently, about one in 11 people in the world are above 65, and this is expected to increase to one in six by 2050. The number of people above 80 years is projected to rise from 143 million in 2019 to 426 million in 2050 (United Nations Department of Economic and Social Affairs Population Division, 2017).

There has been extensive research into biological and neurological alterations with aging. It is well-known that aging causes a decline in processing speed, working memory and inhibitory function, as well as atrophy in several brain structures (Park and Reuter-Lorenz, 2009; Brookmeyer et al., 2007). These normally-appearing damages intensify with aging-related diseases, making the discrimination between normal and disease-related aging both challenging and important (Reuter--Lorenz and Lustig, 2005).

Alzheimer's disease is a chronic neurodegenerative disease causing the death of neurons. As neurons commonly do not reproduce or get replaced, preventing damage in the first place is crucial to slow its progression. There is no cure for AD – even moderate forms refuse treatment – but medication can affect patients with mild forms of the disease (Dodel et al., 2013; Montgomery et al., 2003). For this reason, as well as optimizing treatment plans, early disease detection and prediction is crucial (Siemers et al., 2016; Guerrero et al., 2016).

In aging, brain atrophy is normal. However, in dementia certain regions of the brain have increased speed of atrophy (Park and Reuter-Lorenz, 2009; Leong et al., 2017; Rodrigue and Raz, 2004; Lundervold et al., 2019; Chandra et al., 2019). While the distinction between the neurodegenerative changes by normal aging and those that characterise AD is not evident, studies have shown that greater shrinkage in specific brain regions is linked to AD (Leong et al., 2017; Raz, 2000; West et al., 1994). For example, hippocampal volume reductions and ventricular expansions show different patterns in healthy aging and in dementia (Thompson et al., 2004), and both can be considered as imaging biomarkers to investigate the rate of brain deterioration (Leong et al., 2017; West et al., 1994; Raz, 2000). The change in the brain has been quantified with different methods and techniques, such as counting neuronal cell loss in brain regions (West et al., 1994) and by calculating the changes in the volume of the brain regions from neuroimaging data (Leong et al., 2017; Raz, 2000).

Such imaging findings, and the uncertainty in the clinical diagnosis of AD, leads to both a need and a potential for further quantitative and indicative imaging biomarkers. In recent years, researchers have constructed a variety of analysis tools and approaches to investigate the aging process in the brain using MRI data, often including machine learning methods (Falahati et al., 2014; Guerrero et al., 2016; Jack et al., 2008; Klöppel et al., 2008; Scheltens et al., 1992; Shi et al., 2009).

While there have been many promising results, there are several limitations in these methods and approaches. For example, an assumption underlying many of the proposed machine learning approaches is that the data instances in follow-up MRI examinations are independent and identically distributed. However, in longitudinal data there are certainly correlations (Falahati et al., 2014; Ngufor et al., 2019; Lei et al., 2017), and using proper longitudinal analysis designs have some important advantages, such as reducing the confounding effect of between-subject variability and making it possible to use non-independent data. Some recent works have taken this into account (Ngufor et al., 2019; Lei et al., 2017; Huang et al., 2016; Zhang et al., 2012; Lim and van der Schaar, 2018), but additional limitations remain. One limitation that the present study aims to overcome is an assumption underlying many other approaches: that all subjects have the same number of measurements, and, even, that the measurements are recorded over the same time interval lengths for the entire sample set. In practice, these assumptions are often invalid, leading other studies to remove instances from their data set (Zhang et al., 2012).

In the present study, we propose a pipeline that is better adapted to such situations. It is a framework based on a combination of mixed effects models (LME) and an ensemble machine learning model (Fig. 2). We used linear time-dependent mixed effects model parameters to derive representative features from the MRI measurements in the predictive machine learning models. Our approach applies to situations where subjects have varying number of MRI examinations, potentially recorded at different scan intervals. It is also possible to include subjects that were examined at a single time-point. The mixed effects modelling is applied to the volumetry of brain regions computed by FreeSurfer v.6.0 (Fischl, 2012), enabling extraction of subject- and region-specific longitudinal volume trajectories (Fig. 1). The instability and fluctuations observed when analysing brain structure volumes from MRI over time, caused by e.g. computational instabilities, noise, hydration status, scanner upgrades, time-of-day at scanning (Trefler et al., 2016) or slight variation in the acquisition protocol, and not changes in the brain parenchyma per se, become less influential by using a LME model (Fig. 1b). This makes the representation of individual volume trajectories more robust, and the prediction of longitudinal group differences more precise (Bernal-Rusiel et al., 2013).

Our results show the ability of this framework to make early prediction of AD, prior to clinical diagnosis, and, to a certain extent, distinguish between cognitively normal (CN) subjects and those who are at risk of MCI. Such a model-based predictive framework, together with assessments of risk factors, could have great potential in monitoring natural progression and to evaluate effect of possible therapeutic interventions. It can also help the clinician in prognostics and advice regarding lifestyle changes and preparing patients for likely life events of neurodegenerative disease. In a related work by the authors (Mofrad et al., 2021) we have demonstrated the proposed framework's ability to incorporate any kind of longitudinal measure, in that case cognitive measures from psychometric testing, and also that the MRI-derived measures provide additional information to the predictive model.

2. Methods

We applied mixed effects models to derive features from longitudinal MRI examination, and used the features in machine learning models aiming at predicting MCI and AD prior to the clinical events. Our approach has two key parts: (i) feature selection, model development and validation, and (ii) model-evaluation. We used data from ADNI for the first part, and a combination of ADNI and AIBL data for the second, making sure no subjects were used for both model training and evaluation of predictive performance. The use of the AIBL data for evaluation ensured that our models were evaluated on an independent data set, sourced from different institutions and subjects than those represented in the training set. This is a crucial part of evaluating predictive models as one can otherwise easily overestimate such models' generalization abilities. Fig. 2 illustrates our framework, further explained in this section.

2.1. Data

Data used in the preparation of this work were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by principal investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD (Gavidia-Bovadilla et al., 2017). We also used data collected by the Australian Imaging, Biomarker & Lifestyle Flagship Study of Ageing (AIBL) database (https://aibl.csiro.au). Launched in 2006, AIBL is the largest study in Australia to discover biomarkers, cognitive characteristics, health and lifestyle factors

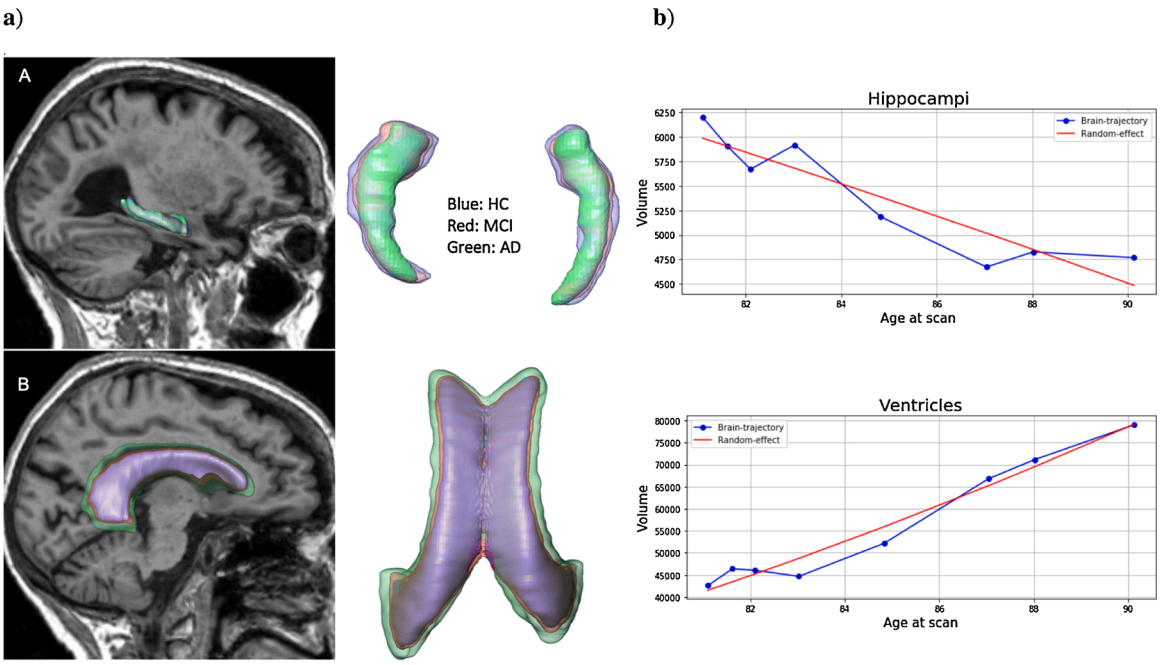

Fig. 1. Aging causes morphometric changes in the brain and dementia accelerate these changes. (**a**) Here we illustrate volume reduction of the hippocampi: left + right hippocampus (A) and expansion of the lateral ventricles (B) with surface renderings from three scans in the series of eight examinations of the same subject. (**b**) A LME model was used to derive representations (i.e. random effects) of such volume trajectories. The blue lines are observed volume trajectories and the red lines are the estimated random effects, based on the eight measurements. Note the small fluctuations or instabilities in the measurements connected by the blue line segments. See the Methods section for more details. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

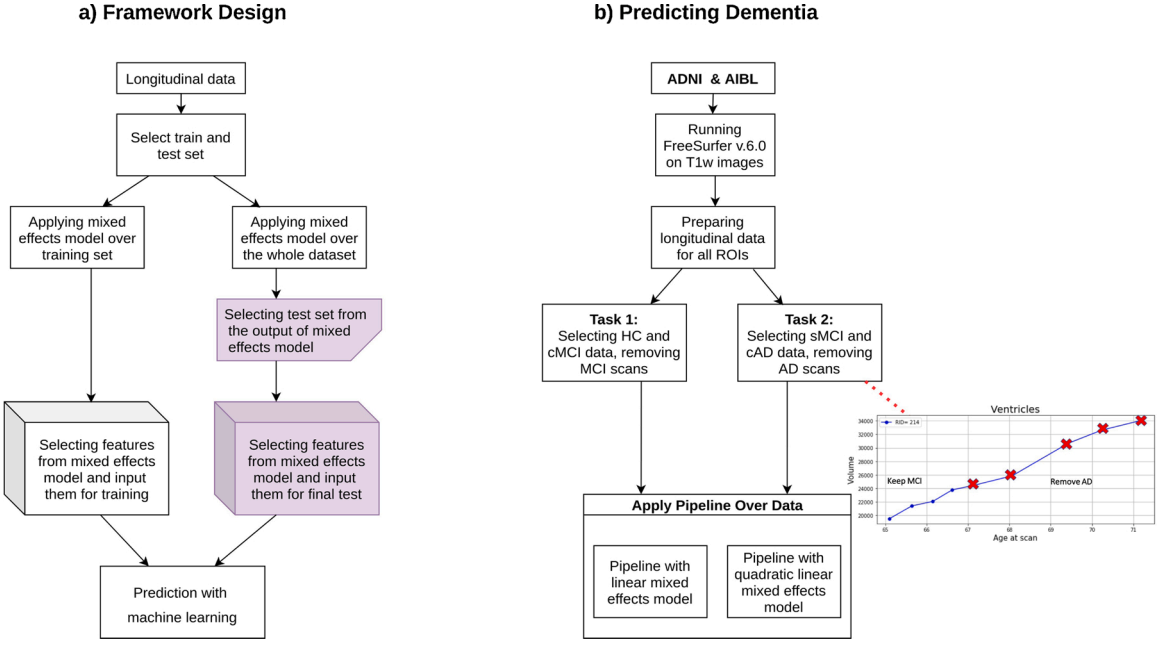

Fig. 2. **a**) Predictive framework for longitudinal data: we first put aside a test set from the longitudinal data. Each mixed effects model is applied separately on the entire data set and on the training set. The features associated with the test set were computed based on constructing the mixed effects models from the entire data set. We used cross-validation on the training set for machine learning model selection. **b**) Prediction of dementia: we first ran FreeSurfer v.6.0 on the longitudinal data from ADNI and AIBL. Then we prepared a table of volumes for brain regions and other information of the subjects. For detecting MCI in task 1, described in Section 3.1, we selected HC and cMCI subjects and removed scans labelled as MCI for all subjects. For detecting AD in task 2, described in Section 3.2, we selected sMCI and cAD subjects and removed scans labelled AD. Finally, a pipeline based on linear time-dependent mixed models was applied.

determining the development of symptomatic AD. It comprises more than 1000 participants with a minimum age of 60 years and contains healthy volunteers, MCI and AD subjects. AIBL study methodology has been reported previously by Ellis et al. (Ellis et al., 2009).

From these cohorts we used longitudinal brain MRI data from subjects scanned multiple times (at least twice) over a period of 15 years. Our data collection consisted of 1673 subjects (with a total of 8002 scans) from ADNI (7764 scans from 1603 subjects) and AIBL (238 scans

from 70 subjects).

2.2. Volumetric biomarkers

The ADNI data release contains derived subcortical and cortical measures computed using FreeSurfer on the T1-weighted MR images. FreeSurfer is a powerful, widely used software package providing automated analyses of structural and functional neuroimaging data from cross-sectional or longitudinal studies (Fischl, 2012).

However, the FreeSurfer data released by ADNI is based on two different software versions, v.4.3 (from March 2009) and v.5.1 (May 2011), both of which are superseded by v.6.0 released in January 2017. Previous studies have demonstrated significant discrepancies between different versions of FreeSurfer (Chepkoech et al., 2016; Gronenschild et al., 2012; Klauschen et al., 2009), and our own exploratory data analyses based on the ADNI data also demonstrate such an effect. For example, Fig. 3a indicates the dissimilarity of volume measurement with the two versions of FreeSurfer the ADNI consortium used on their data, v.4.3 and v.5.1. The results in Fig. 3 show clear discrepancies between hippocampus volumes derived from scanners of field strengths 1.5 Tesla and 3.0 Tesla. This highlights the importance of not changing the version of FreeSurfer during longitudinal studies, especially those involving scanners of different field strengths.

To get more precise information about the potential negative impact of the varying FreeSurfer versions, we conducted an experiment using FreeSurfer v.5.3 and v.6.0 (the newest version at the time of experiment). We selected 80 subjects, controlling for disease status (CN/Dementia, 40/40), gender (F/M, 40/40), scanner field strength (1.5T/3T, 40/40), and age ([75,80)/[80,85), 40/40). One of the results is shown in Fig. 4, indicates that the effect of FreeSurfer versions differs between CN and Dementia subjects. We concluded that the FreeSurfer version is an important factor for studying atrophy, especially in the small regions of the brain (e.g. the hippocampus), and therefore reprocessed all the ADNI and AIBL data using FreeSurfer v.6.0 on Ubuntu 18.04 GNU/Linux workstations. This gave us the data set used in the remainder of this work.

2.3. Mixed effects models

Our framework is based on linear mixed effects models (LME), a well-established approach to longitudinal data analysis, used to derive regression models from dependent data. In contrast to simpler linear models, LME provides a combination of fixed and random effects as predictor variables (Bell and Jones, 2015; Harrison et al., 2018; Lindstrom and Bates, 1990; Müller et al., 2013; West et al., 2014). Mixed effects models allow the collection of relatively simple, robust, noise-free, and subject-specific representations of brain change over time, as illustrated by the red lines in Fig. 1b, based on age at scan as the covariate.

As some brain ROI volumes versus time show linear cohort behavior while others behave nonlinearly (cf. Fig. 5), we were motivated to use LME models with both linear and nonlinear (quadratic) covariates. Our models are based on the model presented by West et al. (West et al., 2014), also used in our previous work (Lundervold et al., 2019):

$$\mathrm{Vol}_{ij}^{r} = \underbrace{\beta_0^r + \beta_1^r \mathrm{Age}_{ij}}_{\text{fixedeffect}} + \underbrace{b_{0i}^r + b_{1i}^r \mathrm{Age}_{ij} + \varepsilon_{ij}^r}_{\text{randomeffect}}, \quad (1)$$

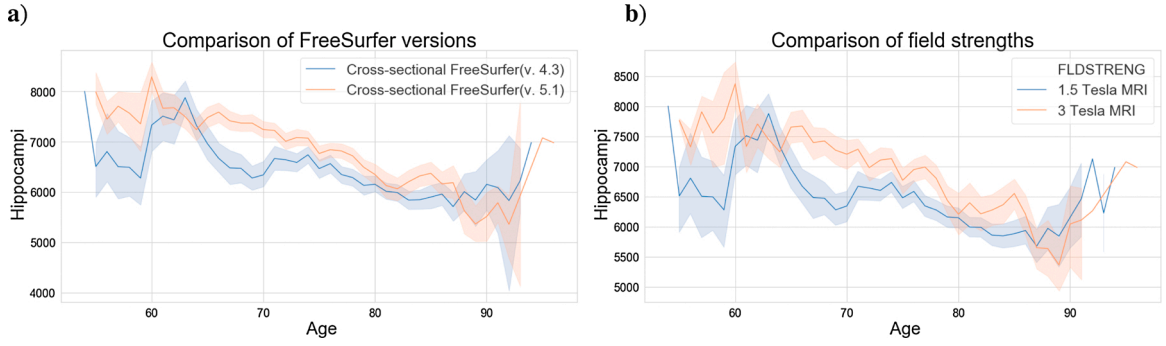

Fig. 3. Plotting the hippocampi volumes for all ADNI subjects across age indicates a discrepancy between (**a**) the volumes calculated by different versions of FreeSurfer and (**b**) the volumes recorded from MRI scanners having different field strengths.

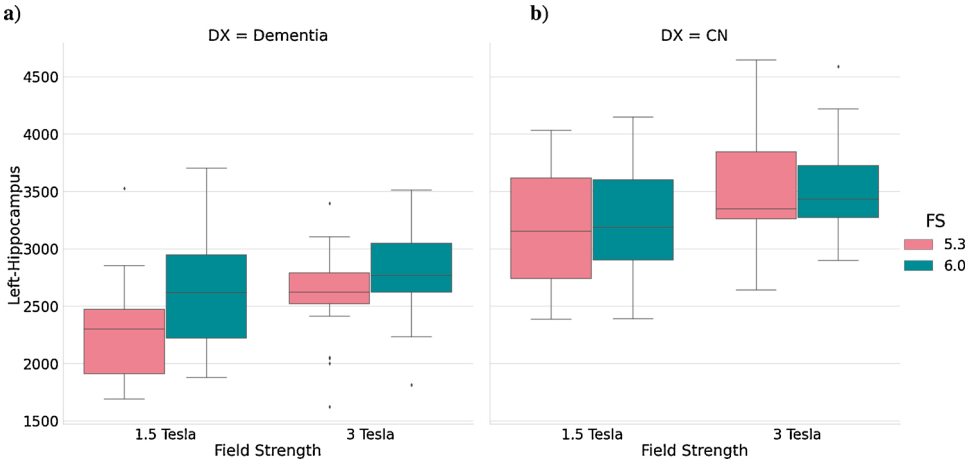

Fig. 4. Box-plots illustrating the importance of FreeSurfer version and magnetic field strength on measuring the volume of the left hippocampus. Each paired box-plot, blue and yellow, contains the same T1w volumes processed with FreeSurfer v. 6.0 and v. 5.3, respectively. **a**) shows volume difference for subjects diagnosed with dementia. **b**) shows volume difference for CN subjects. For dementia the volume discrepancies between FreeSurfer versions are both large and statistically significant (paired t-test, $p < 0.05$) for both 1.5 and 3 Tesla scanners. For CN the version-related differences are insignificant. Note that while we have controlled the gender and age in these groups (1.5 and 3 Tesla) the subjects are different, which makes a precise conclusion of the impact of varying scaner versions difficult. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

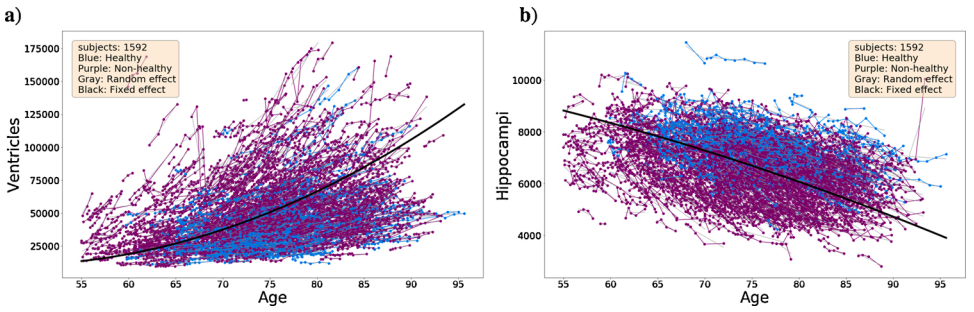

Fig. 5. Longitudinal trajectories for the eTIV-normalised volumes of ventricles (**a**) and hippocampi (**b**) versus age at scan. Healthy subjects are marked in blue, non-healthy in purple. The thick black curve is the cohort fixed effect regression line. The random effects computed by Eq. (2) are shown as thin grey lines for each subject. Note the steady decrease in the hippocampal volumes with time in the age range 55–95 years, and the concomitant nonlinear increase in the lateral ventricle volumes. The plot indicates that the most extensive tissue losses are found among subjects labelled as not cognitively normal. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

where r denotes the brain region, Vol_{ij}^{r} is the volume of region r for subject $i: 1,\ldots,N$ at scan $j: 1,\ldots,n_i$. In our case, n_i varies between 2 and 11. Age_{ij} is age (in years) of subject i at scan j. This is the only predictor variable in the model. The β_0^r and β_1^r are fixed effect parameters, while b_{0i}^r and b_{1i}^r are random effects parameters and ε_{ij}^r denotes the random residual errors.

As seen in Fig. 5, the cohort volume change in the lateral ventricles demonstrate a quadratic behavior, likely due to atrophy over time in multiple brain regions leading to an enlargement of the cerebrospinal fluid-filled lateral ventricles, compensating the tissue loss (i.e. total intracranial volume is preserved). To model this behaviour, we assume the rates of volume change are covariant with both `age` and `age`^{2}. Accordingly, our mixed effect models are:

$$\mathrm{Vol}_{ij}^{r} = \underbrace{\beta_0^r + \beta_1^r \mathrm{Age}_{ij} + \beta_2^r \mathrm{Age}_{ij}^2}_{\text{fixedeffect}} + \underbrace{b_{0i}^r + b_{1i}^r \mathrm{Age}_{ij} + b_{2i}^r \mathrm{Age}_{ij}^2 + \varepsilon_{ij}^r}_{\text{randomeffect}}, \quad (2)$$

where $(\beta_0^r, \beta_1^r, \beta_2^r)$ are fixed effect parameters and $(b_{0i}^r, b_{1i}^r, b_{2i}^r)$ are random effect parameters.

We used the `mixedlm` function in the Python `statsmodels` library (Seabold and Perktold, 2010) (version 0.11.0) to construct and fit the LME models to the data, extracting a mean cohort trajectory (fixed effect) and the subject-specific trajectories (random effects). In this setting the model is linear in the parameters, but can be nonlinear in the covariates. The model parameters (βs and bs) were estimated and stored for each subject, according to Eq. (1) and Eq. (2).

Fig. 5 shows fixed and random effects regressions computed by Eq. (2) for subjects split into two groups: *healthy* (HC, n = 407, f/m = 215/192) and *non-healthy* (sMCI, cAD and sAD, n = 1185, f/m = 492/693). It shows a difference between normal age-related atrophy (blue) and increased atrophy in the case of neurodegenerative disease (purple). This figure indicates the potential of our approach of deriving features from LME models for classifying our different subgroups defined in Table 1a.

We used the volume increase of the ventricles as a measure of total brain atrophy and the volume change in the hippocampi, as it is a well-known structure affected by dementia.

Derived features

From the mixed effects models we derived four features for each individual ROI trajectory: (i) For the linear models (Eq. (1)), a vector of random effect covariates, (b_{0i}^r, b_{1i}^r), containing the intercept of the group and the slope of the random effects line. For the nonlinear models (Eq. (2)) we used the vector $(b_{0i}^r, b_{1i}^r, b_{2i}^r)$, the intercept for the group and the coefficients of `age` and `age`^{2}; (ii) and (iii) The deviation measured at the first scan, d_i^0, and at the last scan, $d_i^{n_i}$, respectively. In other words, the derived random effects values at the first and the last scans, (illustrated in Fig. 6), as given by

Table 1
a) The original ADNI class labels and the longitudinal labels used in the present study with their descriptions. **b)** Total number of subjects and number of T1-weighted MR images according to class label in our study, selected from ADNI and AIBL.

a)

Source	Class	Class description
ADNI	CN	Cognitively normal at visit
	MCI	Mild cognitive impairment at visit
	Dementia	Alzheimer's disease at visit
Our study	HC	CN at all visits
	cMCI	Initially CN, later converted to MCI
	rHC	Risky CN: cMCI with MCI scans removed
	sMCI	MCI at all visits
	cAD	Initially MCI, later converted to Dementia
	rMCI	Risky MCI: cAD with Dementia scans removed
	sAD	Dementia at all visits

b)

Class	ADNI		AIBL	
	ID	#Images	ID	#Images
HC	407	1994	24	90
cMCI	109	596	24	80
sMCI	509	2500	11	34
cAD	269	1540	11	34
sAD	298	1055	-	-
ALL	1603	7764	70	238

$$\mathrm{d}_{\mathrm{i}}^{\mathrm{j}} = \mathrm{Vol}_{ij} - (\beta_0 + \beta_1 \mathrm{Age}_{ij}) \quad (3)$$

and, for the nonlinear models, d_i^0 and $d_i^{n_i}$ given by

$$\mathrm{d}_{\mathrm{i}}^{\mathrm{j}} = \mathrm{Vol}_{ij} - (\beta_0 + \beta_1 \mathrm{Age}_{ij} + \beta_2 \mathrm{Age}_{ij}^2); \quad (4)$$

where j is either 0 or n_i. (iv) The difference of volumes at the first and last scans, divided by the number of years between them (Eq. (5), i.e. the slope of atrophy from the first to the last measurement):

$$\mathrm{Atrophy}_{\mathrm{i}}^{\mathrm{slope}} = \frac{\mathrm{V}_{\mathrm{in_i}} - \mathrm{V}_{\mathrm{i0}}}{\mathrm{Age}_{\mathrm{in}_i} - \mathrm{Age}_{i0}} \quad (5)$$

where V_{i0} and $\mathrm{V}_{\mathrm{in_i}}$ are the volumes at the first and last scans for subject i, respectively. Feature (iv) is motivated by the varying number and timing of scans for the subjects, and that the atrophy seen over e.g. 10 years for one subject can be equal to the atrophy in two years for another (see Fig. 6).

2.4. Predictive models

As input features to our machine learning models we used the subjects' sex, average age at scans, age at last scan, and the above four features from mixed effects models, scaled according to

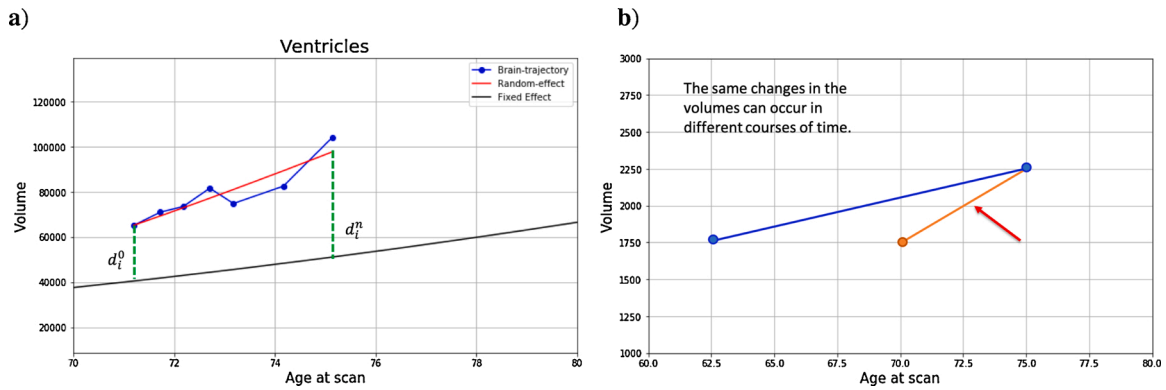

Fig. 6. **a**) Estimated values for random effects at first and last scans (d_i^0, d_i^n) are considered as features (ii) and (iii), respectively. **b**) The linear slop of atrophy (Eq. (5)) calculated based on volumes at first and last scans. The time points are different, and therefore the amount of atrophy in 10 years for a subject can be the same as a 5 years atrophy for another subject (see the red arrow). Therefore, the slope of total atrophy in each ROI is considered as a feature, (iv), for each subject.

$$\text{standardscaling}: \quad \tilde{\mathbf{x}} = \frac{\mathbf{x} - \bar{\mathbf{x}}}{\sigma}, \quad \text{ormax} - \text{minscaling}: \quad \tilde{\mathbf{x}} = \frac{\mathbf{x} - \min(\mathbf{x})}{\max(\mathbf{x}) - \min(\mathbf{x})},$$

where x is a vector of features, $\bar{\mathrm{x}}$ is the mean value of vector x, and σ is the standard deviation for x. We trained an ensemble of a logistic regression and a support vector machine, based on a soft voting strategy, i.e. using the weighted average probabilities from each model in the ensemble. Rather than using single models, with their own specific decision logic, an ensemble constructed from multiple diverse, individually-tuned models can result in a more robust, higher-performing model (Dietterich, 2000; Saeys et al., 2008). We used recall and accuracy scores to assess our models during development and hyper-parameter selection, using subject-level cross-validation on the training set. For each model we set up a grid search through sets of hyperparameters, attempting to find the models with the best generalization abilities.

For the support vector classification model (SVC) we evaluated the regularization parameter C, polynomial, sigmoid and radial basis function kernels, and the kernel coefficient. In `scikit-learn`, the kernel coefficient γ is either set to *scale* or *auto*. For training data with length n, the *scale* setting means that the model uses $1/(\#\text{features} \times \text{variance}(\mathbf{x}))$ as the value of γ and *auto* means it uses $1/(\#\text{features})$. For the logistic regression model we evaluated whether to include an l_2 penalty and the strength of this regularization (C). For both SVC and logistic regression we fixed a random seed, to ensure reproducibility, and set the maximum number of iterations to 500.

We performed feature selection and model development using T1-weighted images from the ADNI dataset, and model evaluation with data from non-overlapping subjects sourced from both ADNI and AIBL. When constructing predictive models for conversion from healthy to MCI and from MCI to AD, we removed all MRI measures taken from the time of conversion and after.

We considered two predictive tasks, described using the subject classes of Table 1:

1 HC subjects (n = 133, f/m = 56/77) versus converted to MCI subjects (cMCI, n = 133, f/m = 55/78),
2 stable MCI subjects (sMCI, n = 279, f/m = 114/165) versus converted to AD subjects (cAD, n = 279, f/m = 111/168).

In task 1 we removed the MRI scans that corresponded to clinical diagnoses of MCI from the cMCI subjects. We call the resulting collection *risky HC* (rHC). In task 2 we removed MRI scans corresponding to AD from the cAD subjects, calling the resulting collection *risky MCI* (rMCI). Details about diagnosis labels and the number of subjects are given in Table 1.

3. Results

3.1. Task 1: Prediction of MCI

We applied our model to two groups of subjects: the subjects marked as cognitively normal at all visits (HC) and the risky HC (rHC, i.e. MRI data from cMCI subjects obtained by removing the scans clinically labelled as MCI). The goal was to investigate whether regular MRI scans can separate HC from rHC, as early detection of MCI based on brain morphometry is an important but also a very challenging task. The subject trajectories for ventricles and hippocampi (Fig. 7) found using LME model show atrophy in the hippocampi and volume increase in the ventricles during aging, while also showing similarity in the trajectories of HC and cMCI. In addition, Fig. 8 shows similar behavior for the average volume of ventricles and hippocampi in HC and cMCI groups of participants, indicating the difficulty of the classification task.

We used data from ADNI for training and data from AIBL for model evaluation. After optimizing the model based on leave-one-out cross validation over the entire training data set (details of hyper-parameters are shown in Table 2 and also in the accompanying code repository[1]), we performed a 15 fold cross validation experiment on the training data set, controlling for labels, age, and gender in the hold-out folds. The mean accuracy and standard deviation obtained by the 15 folds for ventricles and hippocampi ranged from 69 ± 6% to 73 ± 7% (see Table 3 for more details). We then applied the model on the main test set from AIBL for evaluation.

We evaluated the model with eight different feature vectors. First, we extracted four sets of features from the ventricles and the hippocampi volumes, using linear and quadratic LME models. Then we applied the ensemble model on each set of features to find the ones with the highest classification performance. We obtained the best accuracy (71%) for quadratic features extracted from hippocampi. See Table 3 for details about these results.

Next, we combined the extracted features of ventricles and hippocampi to see whether this would improve the classification. The results are shown in Table 3.

3.2. Task 2: Prediction of AD

The ability to predict AD before the symptoms are caught by the clinician is the main objective for our study. We selected sMCI and cAD subjects from ADNI and AIBL to investigate to what extent the atrophy trajectories can distinguish the stable MCI from the risky MCI (subjects

[1] https://github.com/MSamane/A-predictive-framework-for-Alzheimers-disease

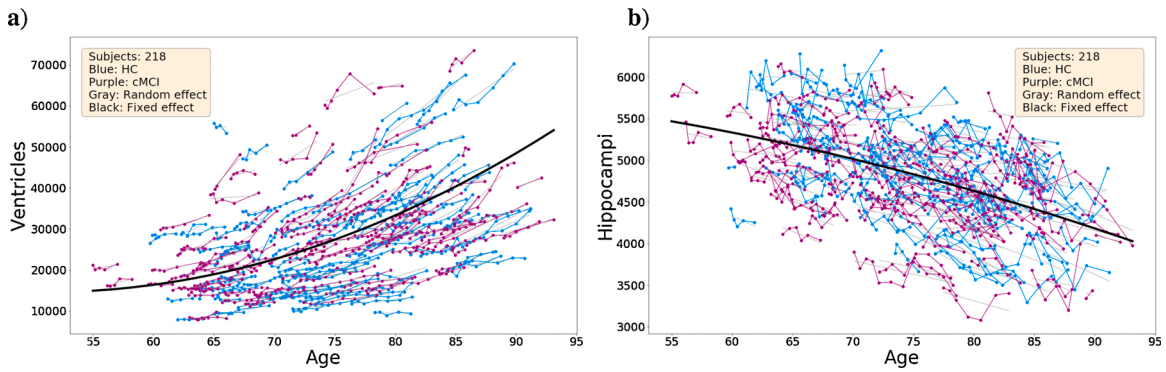

Fig. 7. Task 1: Longitudinal trajectories for the eTIV-normalised volumes of the lateral ventricles (**a**) and hippocampi (**b**) versus age at scan. The thick black curve is the cohort nonlinear regression line. The random effects computed by Eq. (2) are shown as thin grey lines for each subject. The volume of the hippocampi decreases over time, likely contributing to the increase in the lateral ventricle volumes.

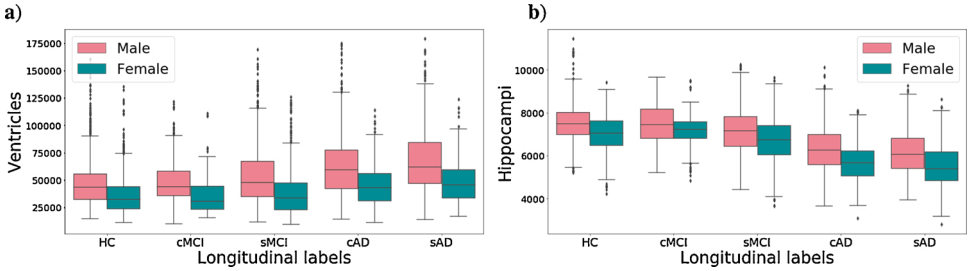

Fig. 8. Ventricles and hippocampi volumes (mm^3) versus our longitudinal subjects from the ADNI dataset, labelled according to Table 1a, show a similarity between HC and cMCI subjects. The sMCI and cAD show a difference in their ventricle volume expansions and their hippocampi atrophy. The difference between ROIs volumes for males and females indicates that gender is an important factor when comparing brain volumes.

Table 2
Model hyperparameters for task 1, for different ROIs-feature combinations, obtained by applying leave-one-out cross validation on the training set.

HC vs. rHC	LME covariates	Logistic regression		SVC		
ROI		scaler	C	scaler	C	kernel
Ventricles	linear	standard	3.13	standard	4.0	poly
	nonlinear	standard	7.78	standard	15.56	poly
Hippocampi	linear	standard	6.7	minmax	7.525	poly
	nonlinear	standard	11.16	standard	10	rbf
Combination	linear	standard	5.6	minmax	6.7	poly
	nonlinear	standard	4.5	minmax	8.9	poly
	nonlin vent, lin hipp	standard	4.5	minmax	20	poly
	lin vent, nonlin hipp	standard	20	minmax	6.7	poly

obtained by removing scans clinically labelled as AD from the cAD subjects). On average, the trajectories for ventricles and hippocampi show a clear atrophy in the hippocampi and a volume increase of the ventricles (Fig. 9). Furthermore, the blue subject trajectories (sMCI) in Fig. 9b are on average above the purple trajectories (cAD).

As for task 1, we first optimized the ensemble machine learning model using leave-one-out cross validation over the entire training data set from ADNI (details of hyper-parameters are in Table 4 and in the accompanying code repository), and then performed a 15 fold cross validation experiment on the training data set, controlling for labels, age, and gender in the hold-out folds, before evaluating the model on a test set. The mean accuracy and standard deviation ranged from 77 ± 4% to 79 ± 6% (see Table 5).

As there are few cAD subjects in AIBL, the test set was constructed using subjects from ADNI (n = 99) and AIBL (n = 22). As for task 1, we applied the ensemble models on eight sets of features extracted by linear

Table 3
Classification results for task 1 for the different ROI features. Note that the accuracy obtained in the 15 fold cross validation experiment is on average better than the accuracy in the final test set sourced from AIBL. As the training and hold-out data in the cross validation are both sourced from ADNI, while the test set is based on AIBL, this is perhaps not surprising.

HC vs. rHC	LME covariates	CrossVal Acc (%)	Accuracy (%)	Precision (%)		Recall (%)		F_1 score (%)	
ROI				HC	rHC	HC	rHC	HC	rHC
Ventricles	linear	69 ± 4	69	65	76	83	54	73	63
	nonlinear	69 ± 6	69	67	71	75	62	71	67
Hippocampi	linear	73 ± 7	67	64	70	75	58	69	64
	nonlinear	71 ± 6	71	67	78	83	58	74	67
Combination	linear	70 ± 7	73	69	79	83	62	75	70
	nonlinear	70 ± 8	65	64	65	67	62	65	64
	nonlin vent, lin hipp	71 ± 8	65	64	65	67	62	65	64
	lin vent, nonlin hipp	72 ± 8	73	69	79	83	62	75	70

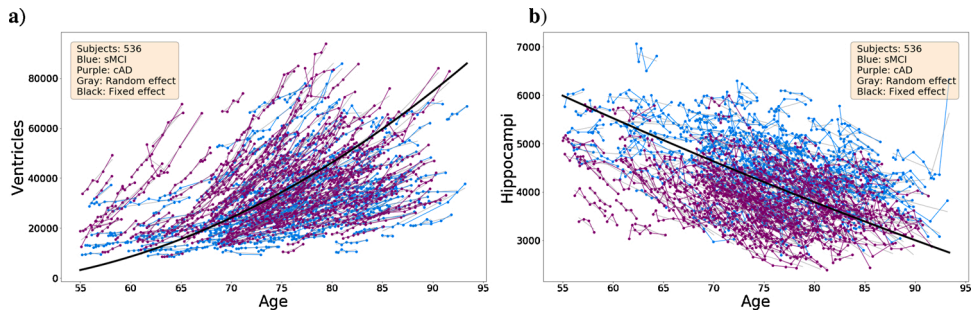

Fig. 9. Task 2: Longitudinal trajectories for the normalised volumes of ventricles (**a**) and hippocampi (**b**) versus age at the scan. The thick black curve is the cohort nonlinear regression line. The random effects computed by Eq. (2) are shown as thin grey lines for each subjects. The volume of hippocampi decreases over time contributing to the increase in ventricular volume. The plot indicates that, the most extensive losses are found among cAD subjects.

Table 4
Model hyper parameters for task 2, for different ROIs-feature combination, obtained by applying leave-one-out cross-validation using the training set.

sMCI vs. rMCI	LME covariates	Logistic regression		SVC		
ROI		scaler	C	scaler	C	kernel
Ventricles	linear	standard	19.9	standard	6.72	rbf
	nonlinear	minmax	8.9	minmax	4.5	poly
Hippocampi	linear	standard	8.89	minmax	11.12	poly
	nonlinear	standard	7.78	minmax	2.23	poly
Combination	linear	standard	6.7	minmax	2.3	poly
	nonlinear	standard	4.5	minmax	5.6	poly
	nonlin vent, lin hipp	standard	7.8	minmax	3.4	poly
	lin vent, nonlin hipp	standard	4.5	minmax	1.2	poly

and quadratic LME from ventricles and hippocampi. The results are presented in Table 5 and in the confusion matrices in Fig. 10 and Fig. 11. The highest accuracy, 78%, was obtained when combining the quadratic features from the hippocampi and ventricles.

4. Discussion

We have developed a flexible and simple framework for extracting features and constructing predictive models from longitudinal MRI in relation to cognitive aging and dementia, based on mixed effects models and ensemble machine learning methods. A strength of the approach is its inherent ability in tackling longitudinal data sets, including situations with sets of subjects with a varying number of scans, taken at different time intervals, which is a common occurrence in longitudinal studies.

We applied the framework to predict dementia, using a large data set sourced from ADNI and AIBL for training and testing. Based on mea-

Table 5
Classification results for task 2, related to different ROI's features. The 15-folds validation results are based on only ADNI dataset (subset of training set) while the other results are based on final test set, a combination of subjects from ADNI and AIBL data.

sMCI vs. rMCI	LME covariates	CrossVal Acc(%)	Accuracy(%)	Precision(%)		Recall(%)		F_1 score(%)	
ROI				sMCI	rMCI	sMCI	rMCI	sMCI	rMCI
Ventricles	linear	77 ± 4	74	77	71	67	80	72	75
	nonlinear	77 ± 4	73	78	69	64	82	70	75
Hippocampi	linear	79± 5	74	78	70	66	82	71	75
	nonlinear	78 ± 5	77	77	77	77	77	77	77
Combination	linear	79 ± 5	75	76	74	74	77	75	75
	nonlinear	79 ± 4.5	78	74	82	85	70	79	76
	nonlin vent, lin hipp	78 ± 5	76	78	75	74	78	76	76
	lin vent, nonlin hipp	79 ± 6	74	73	76	79	70	76	73

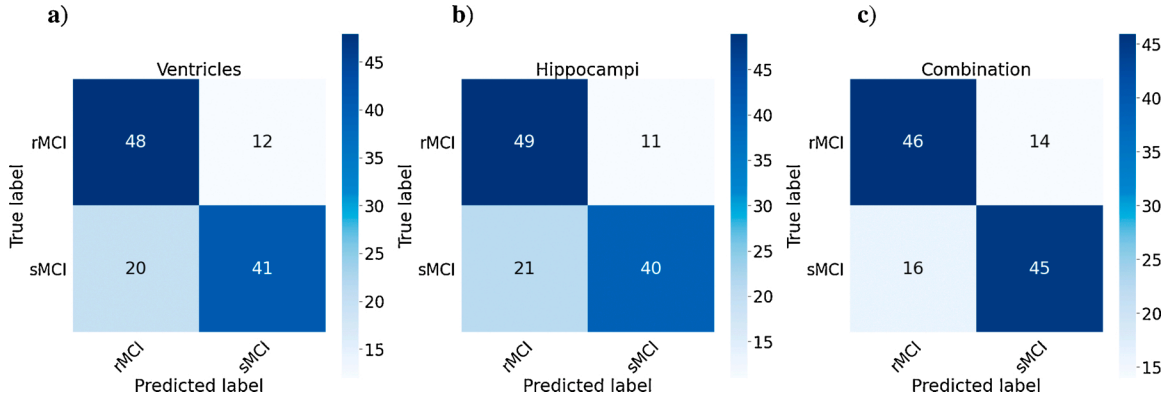

Fig. 10. Confusion matrices for classification of sMCI vs. rMCI based on features extracted from LME model (Eq. (1)) for the ventricles (**a**), the hippocampi (**b**), and the combination of ventricles and hippocampi (**c**).

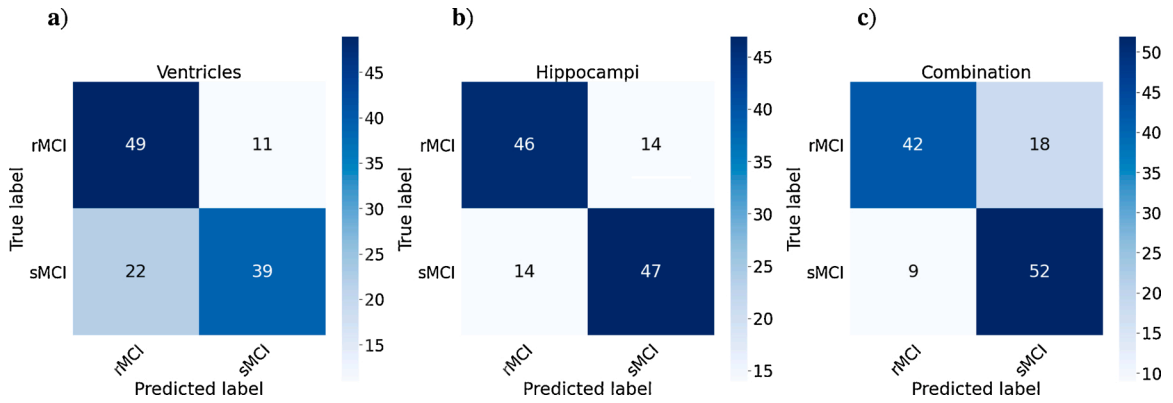

Fig. 11. Confusion matrices for classification of sMCI vs. rMCI based on features extracted from quadratic mixed effects model (Eq. (2)) for the ventricles (**a**), the hippocampi (**b**), and the combination of ventricles and hippocampi (**c**).

surements of hippocampal and lateral ventricle volumes in single subjects over time, we were able to make predictions of conversion from cognitively normal (CN) to mild cognitive impairment (MCI) and from stable MCI to AD, ahead of the corresponding clinical diagnoses, with accuracies of 73% and 78%, respectively. The task of predicting conversion from healthy to MCI is inherently difficult, as it is very challenging to differentiate cognitive decline related to MCI symptoms from cognitive decline with stable cognitive performance at the baseline (Yue et al., 2021). Therefore, our above chance level results at this task is notable. Since the subjects in our study vary with respect to the number of MRI scans and number of years between scans, it is not straightforward to state how early we can predict the risk of MCI or AD prior to diagnosis. In our sample the average time interval between the MRI scans for each subject is 0.53 years (HC: 0.58, cMCI: 0.62, sMCI: 0.48, cAD: 0.53). Therefore, we cannot expect to obtain predictions of conversion to MCI or AD earlier than half a year ahead of the actual conversion.

There are a few studies that predict the conversion from HC to MCI using multi-domain features, including MRI scans (Mofrad et al., 2021; Yue et al., 2021; Albert et al., 2018). Albert et al. (2018) employed imaging-biomarkers related to the hippocampus and the entorhinal cortex in a sample of 224 subjects (178 HC vs. 46 cMCI) obtaining a sensitivity of 64% in predicting the conversion to MCI. Yue et al. (2021) obtained an accuracy/sensitivity of 63%/42% in predicting decline to MCI using MRI-derived features only, improving their results to 70% accuracy and 63% sensitivity when incorporating multi-domain features. Regarding conversion from MCI to AD, Young et al. (2013) predicted this conversion within three years with a 74% accuracy using a Gaussian process classification. This is on par with our results of 78% accuracy. Interestingly, using a deep learning approach (CNN and RNN) and longitudinal MRI data Cui et al. (2019) obtained 72% classification accuracy and 76% sensitivity in their experiments to predict pMCI vs. sMCI.

There are several limitations related to the available data material in our study and in our methods. For example, the group of patients with MCI is highly diverse (Cole and Franke, 2017; Walhovd et al., 2014; Nyberg and Pudas, 2019), and a clinical diagnosis of Alzheimer's disease is inherently uncertain, as the disease is only definite post-mortem (Association, 2013). This is not captured by the labels in ADNI and AIBL, and also holds for similar studies mentioned above.

Furthermore, variability of non-biological origin in MRI measurements, occuring between subjects and in subject examinations over time, will take place (different scanners, calibration issues and scanner drift, different head positions, head motion during scan, etc.) (Trefler et al., 2016; Di et al., 2019). There are also instabilities and uncertainties in the algorithms, libraries and numerical schemes used to compute brain region-specific measures that will lead to sources of variation affecting predictive models and their performance. In this context, we used FreeSurfer v.6.0 and v.5.3 to compute the volumes of the hippocampi and lateral ventricles, exploring some of the inherent variation when using different version of the software and when the images are recorded on scanners of different magnetic field strength (Fig. 3 and Fig. 4). Based on this exploration, we re-computed the volumes in the ADNI and AIBL data sets using the same version of FreeSurfer (v.6.0) to reduce some of the variability. But some instability issues surely remain.

In this work we focused on establishing a framework using only MRI-based morphometric measurements of the hippocampi, as a brain region well-known to be impacted by dementia (Leong et al., 2017; Chandra et al., 2019; Rodrigue and Raz, 2004; Raz, 2000), and the lateral ventricles, as a global measure (proxy) of brain atrophy (Leong et al., 2017; Chandra et al., 2019). Other regions are also impacted by aging and dementia, and inclusions of measures from those ROIs could potentially lead to improved predictions (Rodrigue and Raz, 2004; Raz, 2000; Leong et al., 2017; Hensel et al., 2005; Poulin et al., 2011).

Another approach taken by some researchers (Cui et al., 2019) is to train convolutional and recurrent neural networks to make predictions directly from subjects' MRI recordings (see e.g. Wen et al., 2020, for an overview). This has the possible advantage of bypassing a lot of careful feature engineering and feature selection with its inherent issues, while still making as accurate or more accurate predictions. But it suffers from the disadvantage of leading to less explainable models (Lundervold and Lundervold, 2019).

A major opportunity and motivation for applying machine learning to neuroimaging examinations in middle aged or elderly subjects that are at risk of cognitive decline, mild cognitive impairment or full blown AD, is the ability to make predictions for single individuals. Such imaging procedures and data analysis will thus support *personalized medicine*, and with detailed quantification of image-derived features in combination with subject-specific information obtained from other sources, one can also aim for *precision medicine*. A contribution of the present work is the design and testing of an expressive and flexible machine learning framework that supports both longitudinal image-derived features as well as cognitive scores (Mofrad et al., 2021), where biochemical measures, genetic profiles and other clinical or laboratory measurements can be included. In the context of the present work and available data in the used data repositories, further improvements could potentially be made by including features from multi-modal MRI, such as functional BOLD MRI (Sperling, 2011; Lajoie et al., 2017) and diffusion MRI (Doan et al., 2017), or the presence of the APOE4 gene variant (Kim et al., 2009; Safieh et al., 2019), or values from CSF analyses (Janelidze et al., 2020). Including results from clinical examinations would also be valuable (Holleran et al., 2020), as the present authors have reported in (Mofrad et al., 2021). Challenges for clinical use include the trade-off between locally available measurement techniques and infrastructure (e.g. scanners and protocols), the need for feasible patient examination times, the quality and management of

model predictions in single individuals, and the consideration of available options for therapy and interventions.

Declaration of interests

None.

Authors' contribution

A.S.L and A.L. conceived the approach, S.A.M. and A.S.L. conceived the experiments, S.A.M. conducted the experiments and analysed the results. All authors reviewed the manuscript.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

Our work was supported by the Trond Mohn Research Foundation, grant number BFS2018TMT07. We thank Hauke Bartsch for several useful comments and inputs throughout our work, and for assisting with Fig. 1.

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: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

World Health Organization, 2019. Dementia (accessed 10.09.20). https://www.who.int/news-room/fact-sheets/detail/dementia.
Prince, M.J., 2015. World Alzheimer Report 2015: The Global Impact of Dementia: An Analysis of Prevalence, Incidence, Cost and Trends. Alzheimer's Disease International.
United Nations Department of Economic and Social Affairs Population Division, 2017. World Population Ageing.
Park, D.C., Reuter-Lorenz, P., 2009. The adaptive brain: aging and neurocognitive scaffolding. Annu. Rev. Psychol. 60, 173–196.
Brookmeyer, R., Johnson, E., Ziegler-Graham, K., Arrighi, H.M., 2007. Forecasting the global burden of Alzheimer's disease. Alzheimer's Dementia 3 (3), 186–191.
Reuter-Lorenz, P.A., Lustig, C., 2005. Brain aging: reorganizing discoveries about the aging mind. Curr. Opin. Neurobiol. 15 (2), 245–251.
Dodel, R., Rominger, A., Bartenstein, P., Barkhof, F., Blennow, K., Förster, S., Winter, Y., Bach, J.-P., Popp, J., Alferink, J., et al., 2013. Intravenous immunoglobulin for treatment of mild-to-moderate Alzheimer's disease: a phase 2, randomised, double-blind, placebo-controlled, dose-finding trial. Lancet Neurol. 12 (3), 233–243.
Montgomery, S.A., Thal, L., Amrein, R., 2003. Meta-analysis of double blind randomized controlled clinical trials of acetyl-L-carnitine versus placebo in the treatment of mild cognitive impairment and mild Alzheimer's disease. Int. Clin. Psychopharmacol. 18 (2), 61–71.
Siemers, E.R., Sundell, K.L., Carlson, C., Case, M., Sethuraman, G., Liu-Seifert, H., Dowsett, S.A., Pontecorvo, M.J., Dean, R.A., Demattos, R., 2016. Phase 3 solanezumab trials: secondary outcomes in mild Alzheimer's disease patients. Alzheimer's Dementia 12 (2), 110–120.
Guerrero, R., Schmidt-Richberg, A., Ledig, C., Tong, T., Wolz, R., Rueckert, D., et al., 2016. Instantiated mixed effects modeling of Alzheimer's disease markers. NeuroImage 142, 113–125.
Leong, R.L., Lo, J.C., Sim, S.K., Zheng, H., Tandi, J., Zhou, J., Chee, M.W., 2017. Longitudinal brain structure and cognitive changes over 8 years in an East Asian cohort. NeuroImage 147, 852–860.
Rodrigue, K.M., Raz, N., 2004. Shrinkage of the entorhinal cortex over five years predicts memory performance in healthy adults. J. Neurosci. 24 (4), 956–963.
Lundervold, A.J., Vik, A., Lundervold, A., 2019. Lateral ventricle volume trajectories predict response inhibition in older age – a longitudinal brain imaging and machine learning approach. PLoS One 14 (4), e0207967.
Chandra, A., Dervenoulas, G., Politis, M., Initiative, A.D.N., et al., 2019. Magnetic resonance imaging in Alzheimer's disease and mild cognitive impairment. J. Neurol. 266 (6), 1293–1302.
Raz, N., 2000. Aging of the brain and its impact on cognitive performance: integration of structural and functional findings. In: Craik, F., Salthouse, T. (Eds.), The Handbook of Aging and Cognition. Lawrence Erlbaum Associates Publishers.
West, M.J., Coleman, P.D., Flood, D.G., Troncoso, J.C., 1994. Differences in the pattern of hippocampal neuronal loss in normal ageing and Alzheimer's disease. The Lancet 344 (8925), 769–772.
Thompson, P.M., Hayashi, K.M., De Zubicaray, G.I., Janke, A.L., Rose, S.E., Semple, J., Hong, M.S., Herman, D.H., Gravano, D., Doddrell, D.M., et al., 2004. Mapping hippocampal and ventricular change in Alzheimer disease. NeuroImage 22 (4), 1754–1766.
Scheltens, P., Leys, D., Barkhof, F., Huglo, D., Weinstein, H., Vermersch, P., Kuiper, M., Steinling, M., Wolters, E.C., Valk, J., 1992. Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. J. Neurol. Neurosurg. Psychiatry 55 (10), 967–972.
Falahati, F., Westman, E., Simmons, A., 2014. Multivariate data analysis and machine learning in Alzheimer's disease with a focus on structural magnetic resonance imaging. J. Alzheimer's Dis. 41 (3), 685–708.
Shi, F., Liu, B., Zhou, Y., Yu, C., Jiang, T., 2009. Hippocampal volume and asymmetry in mild cognitive impairment and Alzheimer's disease: meta-analyses of MRI studies. Hippocampus 19 (11), 1055–1064.
Jack Jr., C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., Whitwell, J.L., Ward, C., et al., 2008. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J. Magn. Reson. Imaging 27 (4), 685–691.
Muller, S., Scealy, J.L., Welsh, A.H., et al., 2013. Model selection in linear mixed models. Stat. Sci. 28 (2), 135–167.
Ngufor, C., Van Houten, H., Caffo, B.S., Shah, N.D., McCoy, R.G., 2019. Mixed effect machine learning: a framework for predicting longitudinal change in hemoglobin A1c. J. Biomed. Inform. 89, 56–67.
Lei, B., Jiang, F., Chen, S., Ni, D., Wang, T., 2017. Longitudinal analysis for disease progression via simultaneous multi-relational temporal-fused learning. Front. Aging Neurosci. 9, 6.
Huang, L., Jin, Y., Gao, Y., Thung, K.-H., Shen, D., et al., 2016. Longitudinal clinical score prediction in Alzheimer's disease with soft-split sparse regression based random forest. Neurobiol. Aging 46, 180–191.
Zhang, D., Shen, D., et al., 2012. Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers. PLoS One 7 (3), e33182.
Lim, B., van der Schaar, M., 2018. Forecasting Disease Trajectories in Alzheimer's Disease Using Deep Learning arXiv preprint arXiv:1807.03159.
Fischl, B., 2012. FreeSurfer. Neuroimage 62 (2), 774–781.
Trefler, A., Sadeghi, N., Thomas, A.G., Pierpaoli, C., Baker, C.I., Thomas, C., 2016. Impact of time-of-day on brain morphometric measures derived from T1-weighted magnetic resonance imaging. NeuroImage 133, 41–52.
Bernal-Rusiel, J.L., Greve, D.N., Reuter, M., Fischl, B., Sabuncu, M.R., et al., 2013. Statistical analysis of longitudinal neuroimage data with linear mixed effects models. NeuroImage 66, 249–260.
Mofrad, S.A., Lundervold, A.J., Vik, A., Lundervold, A.S., 2021. Cognitive and MRI trajectories for prediction of Alzheimer's disease. Sci. Rep. 11 (1), 1–10.
Gavidia-Bovadilla, G., Kanaan-Izquierdo, S., Mataró-Serrat, M., Perera-Lluna, A., et al., 2017. Early prediction of Alzheimer's disease using null longitudinal model-based classifiers. PLoS One 12 (1), e0168011.
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., et al., 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. Int. Psychogeriatr. 21 (4), 672–687.
Chepkoech, J.-L., Walhovd, K.B., Grydeland, H., Fjell, A.M., et al., 2016. Effects of change in FreeSurfer version on classification accuracy of patients with Alzheimer's disease and mild cognitive impairment. Hum. Brain Mapp. 37 (5), 1831–1841.
Gronenschild, E.H., Habets, P., Jacobs, H.I., Mengelers, R., Rozendaal, N., Van Os, J., Marcelis, M., 2012. The effects of FreeSurfer version, workstation type, and Macintosh operating system version on anatomical volume and cortical thickness measurements. PLoS One 7 (6), e38234.
Klauschen, F., Goldman, A., Barra, V., Meyer-Lindenberg, A., Lundervold, A., 2009. Evaluation of automated brain MR image segmentation and volumetry methods. Hum. Brain Mapp. 30 (4), 1310–1327.
Lindstrom, M.J., Bates, D.M., 1990. Nonlinear mixed effects models for repeated measures data. Biometrics 673–687.

Harrison, X.A., Donaldson, L., Correa-Cano, M.E., Evans, J., Fisher, D.N., Goodwin, C.E., Robinson, B.S., Hodgson, D.J., Inger, R., 2018. A brief introduction to mixed effects modelling and multi-model inference in ecology. PeerJ 6, e4794.
West, B.T., Welch, K.B., Galecki, A.T., 2014. Linear Mixed Models: A Practical Guide Using Statistical Software. Chapman and Hall/CRC.
Bell, A., Jones, K., 2015. Age, period and cohort processes in longitudinal and life course analysis: a multilevel perspective. In: Burton-Jeangros, C., Cullati, S., Sacker, A., Blane, D. (Eds.), A Life Course Perspective on Health Trajectories and Transitions. Springer, Cham, pp. 197–213.
Seabold, S., Perktold, J., 2010. Statsmodels: econometric and statistical modeling with python. 9th Python in Science Conference.
Dietterich, T.G., 2000. Ensemble methods in machine learning. International Workshop on Multiple Classifier Systems 1–15.
Saeys, Y., Abeel, T., Van de Peer, Y., 2008. Robust feature selection using ensemble feature selection techniques. Joint European Conference on Machine Learning and Knowledge Discovery in Databases 313–325.
Yue, L., Hu, D., Zhang, H., Wen, J., Wu, Y., Li, W., Sun, L., Li, X., Wang, J., Li, G., et al., 2021. Prediction of 7-year's conversion from subjective cognitive decline to mild cognitive impairment. Hum. Brain Mapp. 42 (1), 192–203.
Albert, M., Zhu, Y., Moghekar, A., Mori, S., Miller, M.I., Soldan, A., Pettigrew, C., Selnes, O., Li, S., Wang, M.-C., 2018. Predicting progression from normal cognition to mild cognitive impairment for individuals at 5 years. Brain 141 (3), 877–887.
Young, J., Modat, M., Cardoso, M.J., Mendelson, A., Cash, D., Ourselin, S., Initiative, A. D.N., et al., 2013. Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. NeuroImage: Clinical 2, 735–745.
Cui, R., Liu, M., Initiative, A.D.N., et al., 2019. RNN-based longitudinal analysis for diagnosis of Alzheimer's disease. Comput. Med. Imaging Graph. 73, 1–10.
Cole, J.H., Franke, K., 2017. Predicting age using neuroimaging: innovative brain ageing biomarkers. Trends Neurosci. 40 (12), 681–690.
Walhovd, K.B., Fjell, A.M., Espeseth, T., 2014 Jun. Cognitive decline and brain pathology in aging-need for a dimensional, lifespan and systems vulnerability view. Scand. J. Psychol. 55, 244–254.
Nyberg, L., Pudas, S., 2019 01. Successful memory aging. Annu. Rev. Psychol. 70, 219–243.
Association, A.P., 2013. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). Pilgrim Press, Washington.
Di, X., Wolfer, M., Kühn, S., Zhang, Z., Biswal, B.B., 2019. Estimations of the weather effects on brain functions using functional MRI-a cautionary tale. bioRxiv 646695.
Hensel, A., Wolf, H., Dieterlen, T., Riedel-Heller, S., Arendt, T., Gertz, H.J., 2005. Morphometry of the amygdala in patients with questionable dementia and mild dementia. J. Neurol. Sci. 238 (1–2), 71–74.
Poulin, S.P., Dautoff, R., Morris, J.C., Barrett, L.F., Dickerson, B.C., et al., 2011. Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity. Psychiatry Res. Neuroimaging 194 (1), 7–13.
Wen, J., Thibeau-Sutre, E., Diaz-Melo, M., Samper-González, J., Routier, A., Bottani, S., Dormont, D., Durrleman, S., Burgos, N., Colliot, O., et al., 2020. Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation. Med. Image Anal. 101694.
Lundervold, A.S., Lundervold, A., 2019. An overview of deep learning in medical imaging focusing on MRI. Zeitschrift fur medizinische Physik 29, 102–127.
Sperling, R., 2011. The potential of functional MRI as a biomarker in early Alzheimer's disease. Neurobiol. Aging 32, S37–S43.
Klöppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N. C., Jack Jr., C.R., Ashburner, J., Frackowiak, R.S., 2008. Automatic classification of MR scans in Alzheimer's disease. Brain 131 (3), 681–689.
Lajoie, I., Nugent, S., Debacker, C., Dyson, K., Tancredi, F.B., Badhwar, A., Belleville, S., Deschaintre, Y., Bellec, P., Doyon, J., et al., 2017. Application of calibrated fMRI in Alzheimer's disease. NeuroImage: Clinical 15, 348–358.
Doan, N.T., Engvig, A., Persson, K., Alnæs, D., Kaufmann, T., Rokicki, J., Córdova-Palomera, A., Moberget, T., Brækhus, A., Barca, M.L., et al., 2017. Dissociable diffusion MRI patterns of white matter microstructure and connectivity in Alzheimer's disease spectrum. Sci. Rep. 7, 45131.
Kim, J., Basak, J.M., Holtzman, D.M., 2009. The role of apolipoprotein E in Alzheimer's disease. Neuron 63 (3), 287–303.
Safieh, M., Korczyn, A.D., Michaelson, D.M., 2019. ApoE4: an emerging therapeutic target for Alzheimer's disease. BMC Med. 17 (1), 1–17.
Janelidze, S., Stomrud, E., Smith, R., Palmqvist, S., Mattsson, N., Airey, D.C., Proctor, N. K., Chai, X., Shcherbinin, S., Sims, J.R., et al., 2020. Cerebrospinal fluid p-tau217 performs better than p-tau181 as a biomarker of Alzheimer's disease. Nat. Commun. 11 (1), 1–12.
Holleran, L., Kelly, S., Alloza, C., Agartz, I., Andreassen, O.A., Arango, C., Banaj, N., Calhoun, V., Cannon, D., Carr, V., et al., 2020. The relationship between white matter microstructure and general cognitive ability in patients with schizophrenia and healthy participants in the ENIGMA consortium. Am. J. Psychiatry pp. appi-ajp.