A survey on machine and statistical learning for longitudinal analysis of neuroimaging data in Alzheimer's Disease

Gerard Martí-Juan[a,*], Gerard Sanroma-Guell[b], Gemma Piella[a]

[a] *BCN Medtech, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain*
[b] *German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany*

Abstract

Background and Objectives: Recently, longitudinal studies of Alzheimer's disease have gathered a substantial amount of neuroimaging data. New methods are needed to successfully leverage and distill meaningful information on the progression of the disease from the deluge of available data. Machine learning has been used successfully for many different tasks, including neuroimaging related problems. In this paper, we review recent statistical and machine learning applications in Alzheimer's disease using longitudinal neuroimaging. **Methods:** We search for papers using longitudinal imaging data, focused on Alzheimer's Disease and published between 2007 and 2019 on four different search engines. **Results:** After the search, we obtain 104 relevant papers. We analyze their approach to typical challenges in longitudinal data analysis, such as missing data and variability in the number and extent of acquisitions. **Conclusions:** Reviewed works show that machine learning methods using longitudinal data have potential for disease progression modelling and computer-aided diagnosis. We compare results and models, and propose future research directions in the field.

Keywords: Longitudinal; disease progression; Alzheimer's disease; machine learning

*Corresponding author
Email address: gerard.marti@upf.edu (Gerard Martí-Juan)

1. Introduction

Alzheimer's disease (AD), the most common form of dementia, is an incurable neurodegenerative disease that affects millions of elderly people worldwide [1, 2]. It is characterized by an accumulation of amyloid-beta (Aβ) proteins in the brain and the formation of tau plaques, which gradually impair cognition, leading to death [3]. Cognitively normal (CN) subjects affected with the disease start to show progressive loss of memory and cognition [4], entering a mild cognitive impairment (MCI) stage, before developing to full-blown AD. Detecting the disease in its early stages is key for a more effective treatment aimed at preventing the degenerative process.

A better understanding of the disease progression is crucial for early diagnosis and personalized therapy. AD is described as a multifactorial disease [5], where several markers represent different pathophysiological processes in the brain, with distinct progression paths. Examples of such markers are brain Aβ deposition, tau injury and neurodegeneration, and they can be used to analyze the disease progression from different perspectives. Appendix A and Appendix B give more details on these markers and their dynamics.

Due to the widespread use and availability of medical devices during the past decades, we now have access to electronic medical records containing a varied set of clinical data coming from multiple sources, including brain imaging scans from different modalities, acquired over time in a longitudinal fashion [6]. Contrary to cross-sectional studies, longitudinal studies allow us to measure the evolution and effects of phenotypic characteristics over time caused by disease progression [7].

Traditional methods of data analysis for extracting knowledge rely on searching relationships among measured quantities (variables). In statistical inference,

we start from a hypothesis of the effect of potential independent (input) variables on the dependent (output or outcome) variables, and look for above-chance associations that confirm or refute the hypothesized relation. Statistical modeling caters for prediction, but predictive accuracy is not its primary goal. From another perspective, machine learning (ML) allows us to accurately model the relationship between input and output that generalize to unseen data. ML is particularly helpful when dealing with complex and unwieldy data, and when the number of input variables is large. One application of ML is prediction (e.g., diagnosis), but such techniques are also useful to find patterns or relationships in the data. Methodologically, the boundary between ML and statistical methods is fuzzy. We refer the reader to [8, 9] for a more in-depth discussion.

Progression of the disease using cross-sectional information has been examined by many ML studies [10, 11]. Fewer have used a sequence of acquisitions and assessed longitudinal changes directly. Although working with longitudinal data can improve our knowledge of the disease [12], adding a temporal dimension entails difficulties and data analysis problems, such as data imbalance or time alignment, that need to be addressed [13, 14] (see Appendix C for more details).

In this survey, we review current machine and statistical learning studies and identify trends for longitudinal medical imaging analysis, current gaps in the literature and possible future directions. We divide the reviewed studies in two large groups: computer-aided diagnosis (CAD) and progression modelling. We exclude purely statistical inference studies, as our focus is on general-purpose learning techniques. We believe such techniques can leverage longitudinal information and give new insight into AD and dementia progression, due to their ability to manage and explore growing volumes of diverse available data in an exploratory hypothesis-free setting.

This paper is organized as follows: in Section 2, we describe the search criteria used to gather the reviewed works. In Section 3 we discuss their data usage,

focusing on the type of longitudinal data and measured markers. Next, we talk about the tasks addressed by the reviewed works: in Section 4 we analyze works that model the progression of the disease, categorizing by the main method used, and detailing advantages and disadvantages of the reviewed approaches. Then, in Section 5, we analyze papers focused on computer-aided diagnosis of the disease, describing both the methods and the data used, as well as their performance. We then discuss specific aspects: in Section 6, we explain how researchers handle certain problems such as temporal alignment and missing data approaches. In Section 7 we discuss about the reproducibility and interpretability of the reviewed works. Finally, based on our analysis, in Section 8 we discuss the overall results, draw conclusions and suggest possible further research paths.

2. Search methods

We reviewed works that 1) focus on AD or dementia, 2) use medical imaging derived markers, 3) use longitudinal data, and 4) use ML methods. We created four groups of keywords for the search:

- **Keywords related to the disease:** Dementia, Alzheimer's disease, Mild Cognitive Impairment, AD
- **Keywords related to markers of the disease:** MRI, PET, medical imaging, Magnetic resonance imaging, Positron emission tomography, fMRI, T1 MRI, T2 MRI, FLAIR, DWI
- **Keywords related to longitudinal data analysis:** longitudinal, spatiotemporal, temporal, long-term, follow-up, progression
- **Keywords related to machine learning and statistical learning methods:** classification, learning, prediction, data-driven, precision medicine, pattern recognition, artificial intelligence, AI, ML

The keywords were chosen to be general terms for each concept but specific enough to discard unrelated papers. We used the search engines of PubMed, ScienceDirect, Scopus, arXiv and bioRxiv. In each website we retrieved papers that had at least one keyword from each group. Depending on the website, its search engine did not allow to use the full search: that was the case in ScienceDirect, Scopus and bioRxiv, where we had to reduce the number of terms. In these cases, we only used the first two words of each category, which we consider to be the most representative.

We excluded papers that were not related to AD or that did not use longitudinal imaging data in their experiments or models. We also excluded papers that did not use a general-purpose learning approach, as mentioned in the introduction. For example, statistical reports whose goal was to test a specific hypothesis.

Search was done over a time period going from January 1st, 2007 to July 31st, 2019. We obtained a total of 1404 different papers. Then, we removed papers that were out of scope (1300). These include papers that did not use learning methods, that used only cross-sectional data, or that focused on other diseases, but had appeared in our search. After that, we removed duplicates from the remaining papers (44), leaving us with a total of 60 papers. Finally, we did a search on the references and citations of the selected papers to include relevant works that could have been missed by our initial search, adding up to **105** selected works. Figure 1 shows a diagram of the selection process.

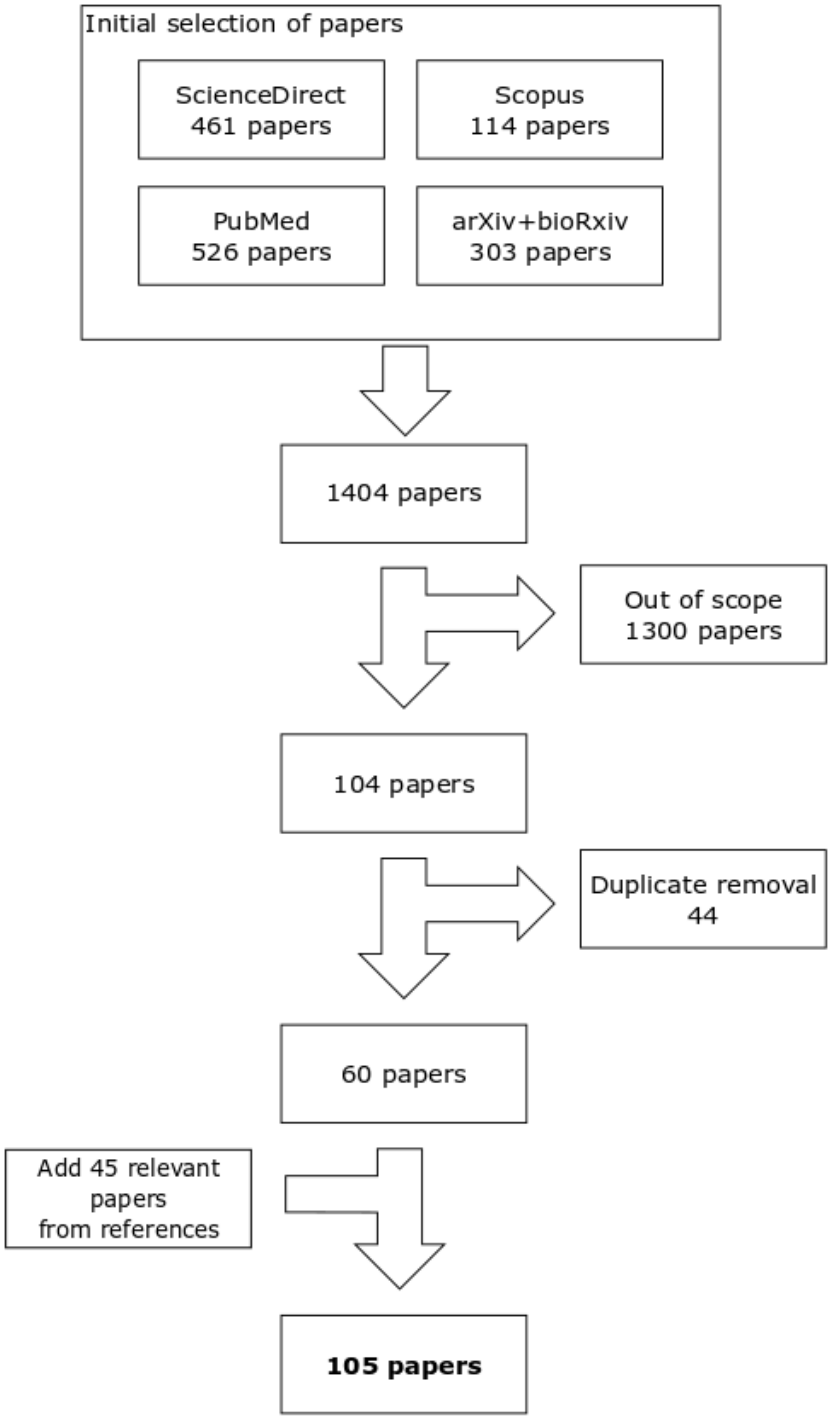

Figure 1: Paper selection pipeline.

3. Data and methods usage

We analyzed the data and methods in the final selection of papers to gain an initial understanding of the reviewed works. We focused on several key aspects: follow-up length, measured markers, database and main methods used.

3.1. Follow-up length

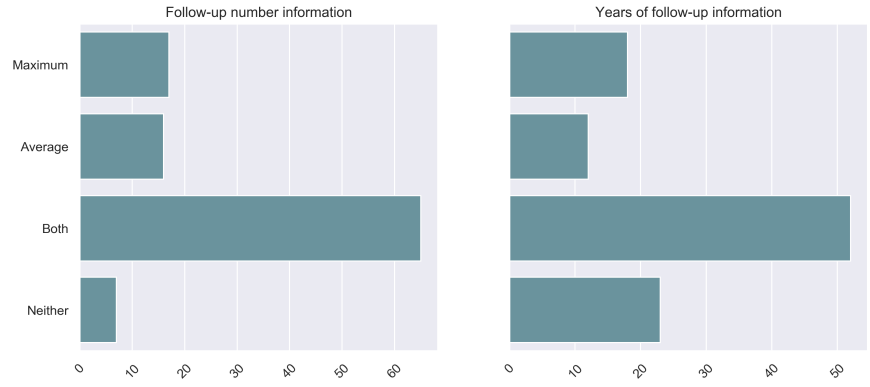

Figure 2: Reviewed papers according to their reporting of follow-up length.

To assess the impact of longitudinal data used in a paper, we analyzed the number of follow-ups per subject and the follow-up length. Figure 2 shows how this information was reported in the reviewed articles. Most of them report enough information on the longitudinal data selection, with the majority of articles reporting both the maximum number of follow-ups and the average number of follow-ups, but a sizeable amount of articles did not report information about the years of follow-up. This can raise some concerns about reproducibility (see Section 7, and it affects the comparison with other works and the correct assessment of the methodology used.

Figure 3 shows the distribution of the (maximum and/or average) number of follow-ups reported in the selected papers. We observe that the distribution is skewed to low number of follow-ups, in both maximum- and average-number

graphs. This is an expected result due to the limited availability of adequate long-term data: subjects having large number of follow ups are sparse in available longitudinal datasets [6]. A substantial percentage of works (39%, not counting those that did not report the information) use only one or two follow-ups in addition to the baseline.

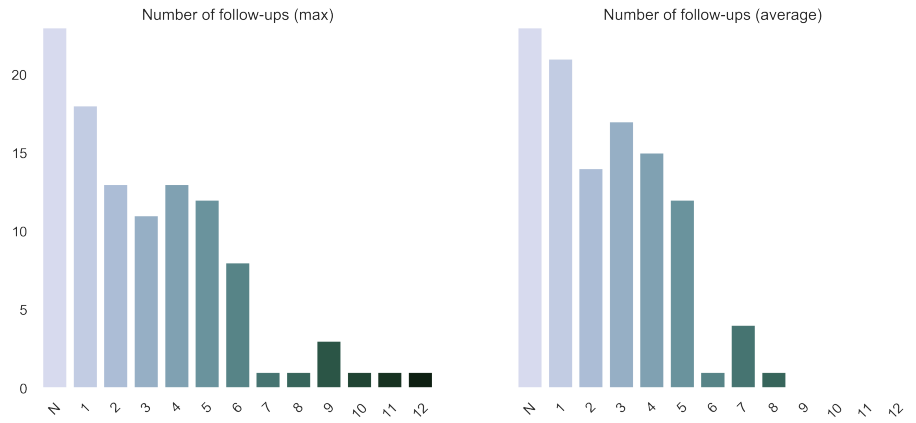

Figure 3: Distribution of the number of follow-ups (maximum and average) used in the reviewed papers. Rounded average values. N: not reported.

Figure 4 shows the distribution of the (maximum and/or average) number of years of follow-up reported. In average, studies follow the patients for between 1 to 3 years, and for a maximum of 2 to 4 years. There is a small subset of papers that are very long term, following the patient from 8 to 12 years [15–20]. Depending on the follow up length, we distinguish between two groups:

- Short-term longitudinal works, including follow-ups up to, at maximum, two years. These papers tend to select a subset of available data, avoiding missing data and unbalanced data problems [21–27] (see Appendix C for an outline of main challenges in longitudinal data). Papers based on short-term brain atrophy [28–32], where usually only one follow-up is needed to calculate tissue loss, and those using several modalities that want to avoid missing data problems [33–35] tend to be short-term.

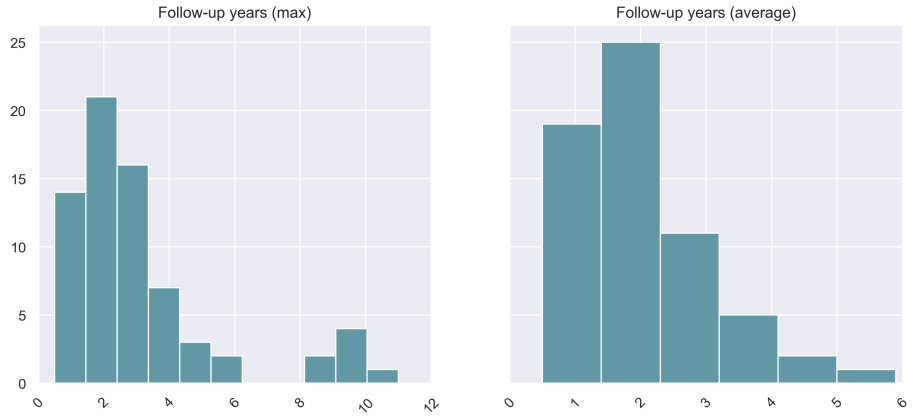

Figure 4: Distribution of the years of follow-up (maximum and average) used in the reviewed papers.

- Long-term longitudinal works, including follow-ups of three years or more. These papers need to deal with missing data, as long-term data tend to be sparse. However, long-term data give an additional insight of the whole progression of the disease that short-term longitudinal data cannot provide. There are many disease progression studies using long-term longitudinal data [17, 19, 20, 36–40], whereas computer-aided diagnosis works using long-term longitudinal data are less common [41, 42]. Some works use longer-term longitudinal data for model validation after training on a short-term subset [43].

As mentioned before, subjects with large number of follow-ups are limited [6]. This is due to various reasons: dropout from the study, missing data due to faulty screenings, and short follow-up time, among others. Even if the number and length of follow-up measures may increase over time, dropout of patients and missing data are phenomena that are present in longitudinal studies. In Section 6 we give examples of methods used by researchers to overcome those problems.

3.2. Data sources

Longitudinal data used in the reviewed papers come from diverse studies. An in-depth review of available longitudinal studies that measure AD markers can be found in [6]. Appendix A gives additional information about studies and initiatives. Figure 5 shows the distribution of databases used in the reviewed papers.

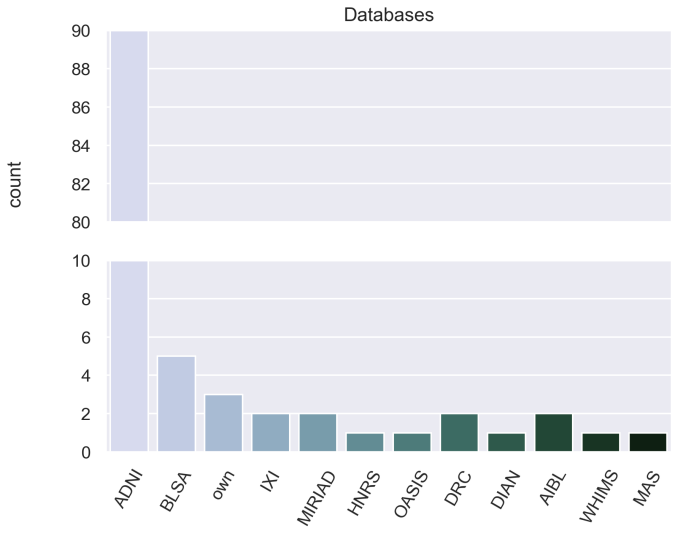

Figure 5: Distribution of longitudinal studies used in the reviewed papers. Graphic is truncated to represent higher counts. ADNI: Alzheimer's Disease Neuroimaging Initiative, BLSA: Baltimore Longitudinal Study of Aging, own: own datasets, not public and/or gathered inhouse, IXI: Information eXtraction from Images, MIRIAD: Minimal Interval Resonance Imaging in Alzheimer's Disease, HNRS: Heinz Nixdorf Recall Study, OASIS: Open Access Series of Imaging Studies, DRC: Dementia Research Centre, DIAN: Dominantly Inherited Alzheimer Network, AIBL: Australian Imaging, Biomarker & Lifestyle Study of Ageing, WHIMS: Women's Health Initiative Memory Study, MAS: Sydney Memory and Aging Study

ADNI [44] is the most used database, being the largest public longitudinal database for AD patients in the world. It has well organized and processed data,

and it has many different modalities and long follow-up times. In addition, obtaining access to ADNI data is easy and fast. With all these characteristics, it is not surprising that ADNI is the most widely published dataset. Other databases are less popular due to their lower amount of data, limited/smaller number of modalities, or difficulty to access them, and often they are used together with ADNI [45–47] as a separate testing set. Some datasets are more specific, regarding, for example, the type of patients they have (Sidney MAS, DIAN, WHIMS) or their follow-up criteria. Interestingly, the Rotterdam elderly study [48] was not used in our reviewed studies, given its size and popularity. However, given that it is not specifically focused on AD and that it has a more complicated access permission, it is reasonable to think that no studies done with this dataset fit our criteria.

Having a predominant database allows for more direct comparisons between results obtained by different methods. However, this can also lead to a generalization problem, where methods would be specific to ADNI's dataset domain but would not extrapolate to the general population. Using an independent (out-of-study) dataset to test the method is advisable to detect this problem, but those are not always available. Data accessibility in medical imaging is a complicated issue due to privacy concerns, which complicates the availability of public datasets.

3.3. Markers

Many works use markers from different sources to characterize the different AD processes. Figure 6 shows the distribution of markers used in the selected papers. We observe that magnetic resonance imaging (MRI) is, by far, the most used type of data. Other modalities, such as positron emission tomography (PET) images with different contrasts or cerebrospinal fluid (CSF) markers, receive low attention despite their importance for disease stratification and early detection [49–51], and are often used in combination with other modalities. Due

to their more invasive acquisition methods, they are not as widely available as MRI in longitudinal studies, and are more prone to missing acquisitions.

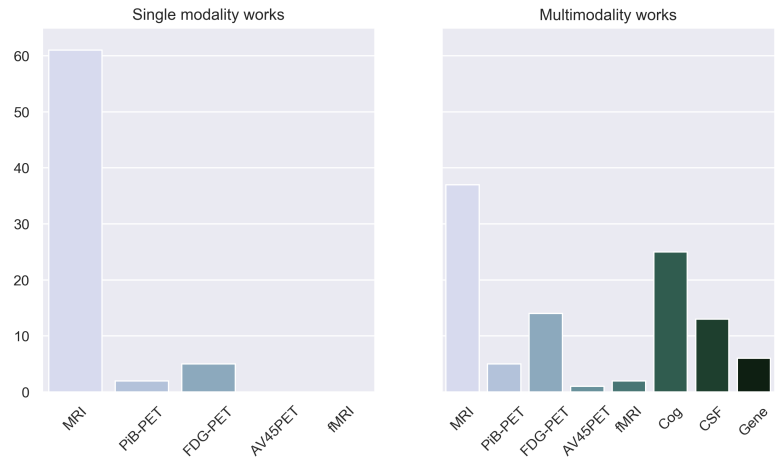

Figure 6: Distribution of measured markers, for single- and multimodality-based papers. MRI: magnetic resonance imaging. PET: positron emission tomography, PiB: Pittsburgh compound B, FDG: fluorodeoxyglucose, AV45: Florbetapir AV-45, fMRI: functional MRI, Cog: cognitive assessments, CSF: cerebrospinal fluid, Gene: genetic markers, Plasma: plasma markers, DTI: diffusion tensor imaging.

3.4. Methods

Table 1 shows the main methods used by the articles reviewed in this paper. Those methods can be categorized into two groups according to the task they aim: progression models, which seek to quantify the evolution of the disease, and classification models, which predicts diagnosis labels of patients. Some of the methods, such as deep learning or multi-task learning, are flexible and can be used for both tasks. Moreover, some works use several different methods for their objectives. The list of methods is not exhaustive, and provides a general picture of the current approaches in the field. For classification, support vector machines (SVM) are the most used method, whereas for progression, a wider

variety of methods are emplyed. In the next sections (Section 4 and 5) we discuss and compare the aforementioned methods.

Method	References
Multi-task learning	[52–58]
Deep Learning	[15, 40, 59–62]
Event-based models	[63, 64]
Manifold learning	[65–67]
Mixed-effect models	[18, 20, 38, 39, 68–72]
Shape analysis models	[46, 73–77]
Gaussian processes	[29, 78–80]
Data-based progression scores	[19, 47, 81–83]
Support Vector Machine	[24, 25, 27, 33, 45, 84–97]
Multiple Kernel learning	[34, 98, 99]
Logistic Regression	[22, 23, 31, 32]
Random Forests	[21, 100, 101]

Table 1: Main methods used in the reviewed papers.

4. Progression models

Models of disease progression can be used to quantify the evolution, determine temporal trajectories and detect different paths of degeneration, among other sub-tasks. In this section we comment on approaches that build disease progression models from longitudinal data, grouping them by their general methodology.

4.1. Multi-task learning for cognitive prediction

Predicting the rate of cognitive decline from imaging markers can be useful to detect brain regions that directly affect cognitive evolution. Also cognitive performance can provide a continuous measure related to disease progression that may complement the categorical information about diagnostic. Since several cognitive scores of the patient are often available, multi-task learning is a popular approach for cognitive prediction. Defining cognitive scores as separate prediction tasks and training them jointly creates a more robust predictive model. Such models have shown to have many advantages: they can use a variable number of follow-ups [55, 58, 90] and can provide direct information between cognitive scores and imaging markers [90, 99, 102–104]. Apart from multi-task learning, other methods have been used for this task, such as probabilistic models [105], regression models [106], or learning ensemble models [41, 100, 107], which combine different, smaller models, and can be defined in flexible ways to integrate missing follow-ups into the model.

4.2. Deep learning

Deep learning is a powerful representation technique that is state of the art in many ML problems. A deep learning model is a neural network with many layers, which is able to learn from large amounts of data to do specific tasks. For a comprehensive review of deep learning techniques in medical imaging, we refer the reader to [108].

Convolutional neural networks (CNN), which are commonly applied with images, have been used for cognitive score prediction [54, 59]. In [54], they combined CNN and MTL, using CNN-based features to train a MTL-based model. In a different approach, [59] proposed a CNN architecture that can predict the cognitive score of the patient at any time, not being restricted to existing follow-ups.

Recurrent neural networks (RNN) are networks where previous outputs are used as inputs while having hidden states. Due to this, they are said to have memory and are able to model sequential data. Consequently, RNN have potential to be able to learn from longitudinal data. They have already been used to predict progression of AD using diverse cognitive scores [109], but without imaging information. [15, 40, 61] all used RNN with imaging markers, although not to quantify progression but for computer-aided diagnosis. Despite their popularity in other fields, RNN are still not widely used for modelling disease progression using medical imaging. A reason could be that deep learning methods are mostly non-interpretable, and any performance gains do not usually compensate the loss in interpretability.

4.3. Event based models

Event based models (EBM) [63] are a modelling approach that describe a neurodegenerative disease by an ordered series of events, such as a new symptom appearing on the patient. EBM define a fixed number of markers, modelling each of them separately to obtain a distribution of abnormal and normal values (indicating the presence of the disease or not) and to generate an ordering of events. We can use this ordered sequence of events to assess the disease stage.

In their initial formulation, EBM were not well-suited to accommodate longitudinal data. There are several studies based on EBM for modelling disease progression that use longitudinal data, either by using markers derived from brain atrophy rate [28, 63, 110] or because they validated their results using the available longitudinal information [64, 110]. Given the potential and strong results they show, integrating longitudinal data in such models is a promising research direction.

4.4. Manifold learning

Manifold learning is an approach to dimensionality reduction. It assumes that the high-dimensional data lie (at least approximately) on a manifold of much lower dimension. Applied to disease progression modelling, manifold learning can be used to estimate an underlying subspace where longitudinal patient trajectories can be better represented. These are directly learnt from the data, and typical problems such as temporal alignment and unbalanced datasets are solved in the subspace learning process. [66] used Laplacian eigenmaps to build a longitudinal manifold, where AD and CN subjects are well differentiated. [65, 67] also proposed a method based on Laplacian eigenmaps, adding a constraint to limit connections between scans from the same subject, to create a temporal embedding that shows the progression of the patient.

These methods are not as popular as other types of models for progression: they are not as directly interpretable as other models such as EBM or multi-task learning, and can be much more complex to implement. However, one can introduce interpretability by effectively embedding the progression in a relevant low dimensional manifold. This, together with their potential to integrate contextual information (e.g. constraints for longitudinal data [67]), make manifold learning an underrated approach for disease progression.

4.5. Mixed effect models

Mixed effect models are widely used due to their flexibility to deal with unbalanced data, and their ability to naturally model average disease progression (fixed effects) and inter-subject variability (random effects). They are commonly considered a classical statistical technique, but we decided to include works that used such models to observe and quantify progression, due to their importance in longitudinal analysis. [111] presented an overview of these methods for MRI, suggesting that linear mixed effect models detect MRI longitudinal group differences with more sensitivity and specificity than other methods such as ANOVA

or general linear models, specially with unbalanced datasets.

Mixed effect models have been extensively used in longitudinal progression models for AD [19, 20, 45, 69–72, 112, 113], as they can be directly adapted to unbalanced longitudinal datasets [69], extended by adding priors such as genetic markers [72] or different progression speeds and disease onsets [71], or used to derive new data-based markers that reflect the progression of the disease [19, 20, 45, 112]. Apart from MRI, linear mixed models have also been applied to other markers, such as PET-based markers [20, 68].

Mixed effect models are easily adapted to more than one modality of data. For this reason, they have also been used extensively in longitudinal multimodal analysis. The easiest way to use them on multiple modalities is to define a different model for each modality [18, 38, 68], allowing us to draw comparisons between markers [82] or to combine them to show the overall progression of the patient [19, 20, 82, 114, 115]. This combination usually needs some kind of temporal alignment of the markers (see Section 6.1 prior to fitting the model [38], by using a defined scale such as cognitive scores [116], or directly in the fitting model [115]. [39] applied a multifactorial mixed model to one of the largest patient cohort in the field, with more than 7700 images and markers from 1100 patients at various stages of the disease, to model and explore the evolution of different markers. Their results suggest that vascular dysregulation might be the earliest factor associated with AD development.

4.6. Shape analysis models

Some methods focus on modelling shape changes of certain brain regions during the disease. This allows capturing subtle variations between or within subjects, which would be lost just by looking at flat imaging markers such as volume or voxel intensities. Current research shows longitudinal changes in shape in key structures of the brain (such as lateral ventricles or hippocampus) that are strongly related to cognitive degeneration [73, 74] and can reveal differ-

ences between groups of patients [75–77]. Mixed effect models are often used for modelling shape changes [73–76], with [46] proposing a novel vertex clustering method to model shape changes over time, using a similar mixed effect model already proposed in other reviewed works [71, 82, 114].

4.7. Other models

Besides the methods discussed so far, other types of models for disease progression are also used. Generalized estimation equations for longitudinal analysis [117] model both the mean response of a population and the covariance of repeated measures [118], and deal with unbalanced datasets [37, 119]. Gaussian processes are also used to model spatio-temporal changes and dependencies [29, 78, 79] and to integrate different modalities [80]. Some methods focus on specific tasks, such as finding hidden latent temporal factors of the disease [16, 120], or are specifically designed to tackle problems such as unbalanced data or patient alignment [119, 121–123]. In [124] they used a non-linear atlas-based model to simulate future MRI scans from previous follow-ups, similar to [79]. Other papers focus on specific data or markers, such as brain connectivity [125] or functional data [17].

[47, 81, 126] defined a method that aggregates the aging change over time on a single marker, and compared it to the real age of the patient and its evolution across time. Atrophy of the brain due to dementia can be similar to the atrophy due to normal aging: brain atrophy can occur because of both normal aging and pathological reasons. Disentangling those two sources can be useful to discover the disease earlier. In the same line, [127] proposed a non-rigid registration method to disentangle the contributions of normal atrophy and disease atrophy, and observed their relation to AD progression. This age-based approach allows us to introduce prior knowledge about healthy subjects in the model and then study the deviations that appear. Such idea has been used in computer-aided diagnosis with remarkable results [89].

For multimodal data, we find works combining genetic information with imaging markers, using regression based models [32, 128], which are useful to discover longitudinal interactions between imaging markers and genetic factors that could go unseen in cross-sectional analysis [12]. Creating data-based progression scores [45, 82, 83, 87, 129] is also a recurrent approach to quantify progression, independently of the method chosen, because such scores are easy to interpret and compare against. [130] applied hidden Markov models, using different markers, to characterize the progression of the disease into stages, and discovered two different paths of AD progression. Recently, [131] proposed a vertex-wise progression model over cortical surface which can be used in different diseases and with different markers to recover and estimate patterns of brain pathology. Their test the model on real and simulated data and is shown to have potential clinical relevancy. Using multimodal data can also be useful to find subtypes of the disease, given its heterogeneity [132]. Thus, combining longitudinal multimodal data is effective to discover hidden AD factors unseen in cross-sectional or single-modal data.

5. Computer-aided diagnosis

Many ML approaches on AD focus on computer-aided diagnosis (CAD): given data about patients, the objective is to classify them depending on their diagnosis (CN, MCI or AD). An extension of this task is to distinguish between MCI patients that will convert to AD (MCIc), and those that will not convert and remain stable (MCInc), which is useful for early detection of the disease. Cross-sectional imaging data have been used extensively for this task, and we refer the reader to [133] for an in-depth review. In this section we study works dealing with this problem using novel methods to process and interpret longitudinal data. We present first the papers using exclusively structural MRI (which are the majority), and then those using other modalities. Finally, we comment on the general performance of the methods.

5.1. Computer-aided diagnosis - structural MRI

A large percentage of the reviewed works use structural MRI as their main source of data, and CAD focused works are no exception, given that structural MRI is considered a clinical predictor of AD [96]. Table 2 shows the performance and principal characteristics of the reviewed works using (exclusively) MRI on CAD. We have divided them by their use of MRI data to build their model: voxel-wise methods and region of interest (ROI) based methods. Some of the works focus only on the hippocampus.

Whole brain voxel-wise analysis allows detecting relations across the whole brain, not being restricted to parcellations. However, this results in high-dimensional input data, which often required feature selection and/or dimensionality reduction. One way to do it is to detect landmarks across the brain and extract features around those marks [88, 94]. Other methods are principal components analysis (PCA) [84, 88, 93], regularization [60] or metric learning [26]. All these methods need to capture the differences/changes in voxels across time, and use them as features. This can be done by computing some difference between the images, such as brain volume changes [88], or deformation maps across follow-ups [45, 92, 93]. In [134], they used a hierarchical regression classifier on longitudinal voxel selection features that solves both problems: selecting the voxels using an individual classifier for each single voxel on the brain, and training the classifiers with the longitudinal data. In a more recent study, [61] used two different neural networks (multi-layer perceptrons and gated recurrent units) to extract spatial and longitudinal features from MRI images.

ROI-based methods use a parcellation of the brain to extract features, limiting their amount and avoiding the curse of dimensionality. Studies use a variety of popular methods, such as support vector machines (SVM) [89, 91, 95, 97], multi-task learning [52], or RNN [15], extending them to account for longitudi-

nal data. For example, [95] defined a stratified SVM method to enforce temporal consistency across follow-ups, and [89] created null models of normal aging using non-demented subjects, and training an SVM with the residuals. Those two methods illustrate how prior knowledge of the dynamics of the disease (non-reversible nature of AD and deviation from normal aging, respectively) can be incorporated into a ML model.

The hippocampus is one of the earliest affected brain regions in AD [3]. Consequently, many works have focused on this region [27, 86], specially for MCIc vs MCInc classification [23, 31]. We observe a variety of methods to extract meaningful features from the hippocampus: patch-based atrophy descriptors [31], longitudinal segmentation methods [112], non-linear metric learning and autoencoders [27], longitudinal deformation [22, 23] and hippocampus volume change [86]. These last two articles also compare different processing methods and their performance to further validate their approach.

Among the reviewed papers, SVM is the most used classification method (see Table 2 due to its simplicity, availability and strong performance. However, deep learning is starting to take off, using CNN [60] and RNN [15, 61].

5.2. Computer-aided diagnosis - other modalities

Some works use other types of modalities, such as fludeoxyglucose (FDG)-PET, which can present early indicators of the disease [135], or use multimodal approaches, with various types of data. Table 3 summarizes the performance and characteristics of such studies.

The importance of FDG-PET based markers in longitudinal analysis of AD is shown in [85], where they tested the predictive capacity on MCI conversion depending on the temporal distance to the conversion using longitudinal FDG-PET images. They showed that, although the performance decreases as the

Table 2: Performance of reviewed computer-aided diagnosis papers using longitudinal structural MRI

Study	Subjects				Scans	Type	Algorithm	Database	Validation	Classification results		
	CN	sMCI	pMCI	AD						AD/CN	MCI/CN	sMCI/pMCI
[92]	-	76	27	-	3	V	SVM	ADNI	LOOCV	-	-	81.5
[96]	83	61	142	83	7	V	SVM	ADNI	BS	-	-	62
[88]	30	-	-	30	3	V	SVM	ADNI	LOOCV	91.7	-	-
[26]	123	121	-	94	2	V	SVM	ADNI	CV	88.4	86.5	-
[134]	-	61	70	-	7	V	LSR	ADNI	CV	-	-	79.4
[84]	148	148	-	-	2	V	SVM	HNRS,OASIS	CV	-	74.3	-
[60]	68	-	-	70	6	V	DL	ADNI	CV	94	-	-
[61]	229	-	-	198	5	V	RNN	ADNI	CV	89.69	-	-
[93]	-	47	63	-	5	V	SVM	ADNI	CV	-	-	92
[94]	207	346	-	154	6	V	SVM	ADNI	CV	88.3	79.02	-
[91]	40	36	39	37	5	ROI	SVM	ADNI	LOOCV	96.1	-	81.7
[52]	-	185	164	-	5	ROI	MTL	ADNI	LOOCV	-	-	71.4
[53]	-	53	60	-	4	ROI	MKL	ADNI	CV	-	-	78.2
[95]	-	81	70	-	5	ROI	SVM	ADNI	CV	-	-	76.5
[89]	215	366	-	166	8	ROI	SVM	ADNI	CV	94.1	83.8	76.7
[15]	742				12	ROI	DL+LDA	ADNI	-	0.9[b]	0.59[b]	0.78[b]
[22]	-	84	19	-	2	H	LGR	ADNI	CV	-	-	0.65/0.62[c]
[23]	-	84	19	-	2	H	LGR	ADNI	CV	-	-	0.46/0.84[c]
[86]	148	95	121	96	3	H	SVM	ADNI	CV	-	0.88[b]	-
[31]	-	100	164	-	2	H	LGR	ADNI	CV	-	-	76.6
[27]	123	121[a]	-	94	2	H	SVM	ADNI	CV	85.9	-	76.7
[112]	137	82	101	77	4	H	LDA	ADNI,MIRIAD	BS	0.947[b]	0.805[b]	-

SVM: Support vector machine, LSR: Least squares regression, LGR: Logistic regression, MTL: Multi task learning, MKL: Multiple kernel learning, DL: Deep learning, LDA: Linear discriminant analysis, V: Voxel-wise, ROI: Region of interest, H: Hippocampus, ADNI: Alzheimer's Disease Neuroimaging Initiative, HNRS: Heinz Nixdorf Recall Study, OASIS: Open Access Series of Imaging Studies. CV: k-fold cross validation. LOOCV: Leave one out cross-validation. BS: Bootstrapping. If the number of scans is variable, the maximum is reported.

[a] MCI subjects

[b] AUC (area under the curve)

[c] Specificity and sensitivity

distance to conversion increases, they were able to track AD progression for up to two years before disease onset. Other studies use region-based [24, 33] and voxel-based [25] analysis of FDG-PET imaging, and show that adding longitudinal information to the problem improves classification accuracy, compared to only using cross-sectional data. Reviewed works show improvements in the longitudinal model with respect to the cross-sectional model. This suggests that, at least for FDG-PET, longitudinal data are important to improve CAD models.

Many classification studies use data coming from various modalities. Comparison studies [34, 99, 101] show that multimodality and longitudinal data improve the performance of the model compared to baseline models using only cross-sectional data or with a lower amount of longitudinal data. There are diverse methods to combine the data, ranging from direct concatenation [40] to more complex methods, such as multiple kernel learning (MKL) [136]. MKL allows us to directly combine different modalities and interpret the resulting model as a weighted combination of kernels. Other methods such as linear regression [42], SVM [137], random forests [21, 101], RNN [40] and MTL [56] have also been used to combine and select features from different modalities of data.

5.3. Performance analysis

Using longitudinal data for CAD leads to better performance in hard problems such as early detection of MCI converters [21, 31, 53, 93], where cross-sectional data could be insufficient to determine whether a patient will progress to AD [138]. Tables 2 and 3 show the performances of the reviewed papers focusing on CAD. Performance is reported using accuracy, unless otherwise stated.

For classification of CN vs AD, [26, 27, 60, 89, 91] reported the highest performances, up to 96% accuracy [91]. For classification of CN vs MCI, a harder problem, [89] showed strong results, with 82% accuracy. The hardest problem is distinguishing between converting and non-converting MCI patients, which is

Table 3: Performance of reviewed computer-aided diagnosis papers using other imaging modalities and data types.

Study	Subjects				Scans	Modality	Algorithm	Database	Validation	Classification results		
	CN	sMCI	pMCI	AD						AD/CN	MCI/CN	sMCI/pMCI
[33]	40	-	-	40	3	FDG-PET	SVM	ADNI	LOOCV	78	-	-
[24]	54	64	53	50	2	FDG-PET	SVM	ADNI	-	88	-	63.1
[25]	66	109[a]	-	48	2	FDG-PET	SVM	ADNI	CV	91.2	70.2	-
[85]	-	56	44	-	5	FDG-PET	SVM	ADNI	CV	-	-	81
[42]	-	100	200	-	4	MRI,Cog	LSR	ADNI	LOOCV	-	-	89.7
[137]	-	65	54	-	3	MRI,Cog	SVM	ADNI	CV	-	-	84.3
[21]	-	78	86	-	2	MRI,Cog	RF	ADNI	OOB	-	-	82.3
[101]	-	85	182	-	2	MRI,Cog	RF	ADNI	OOB,CV	-	-	80.2
[34]	66	119[a]	-	48	2	MRI,FDG-PET,CSF,Cog	MKL	ADNI	CV	92.4	-	0.76 [b]
[99]	-	50	38	-	5	MRI,FDG-PET	MKL	ADNI	LOOCV	-	-	78.4
[98]	213				6	MRI,Gen,Cog	MKL	ADNI	-	90[c]	-	-
[56]	-	53	65	-	3	MRI,FDG-PET,Cog	MTL	ADNI	CV	-	-	84
[57]	23	24[a]	-	-	6	MRI,fMRI	MTL	ADNI	LOOCV	95[c]	-	-
[40]	521	864[a]	-	336	23	MRI,PET,CSF,Cog	RNN	ADNI	-	95.8	77.3	85.8[c]

SVM: Support vector machine, LSR: Least squares regression, RF: Random forest, MKL: Multiple kernel learning, MTL: Multi task learning, FDG-PET: Fludeoxyglucose positron emission tomography, fMRI: functional MRI, Cog: Cognitive scores, CSF: Cerebrospinal fluid, Gen: Genetic information. CV: k-fold cross validation. LOOCV: Leave one out cross-validation. BS: Bootstrapping, OOB: Out of bag estimation.

[a] MCI subjects

[b] AUC (area under the curve)

[c] MCI vs AD performance

also the task that has gathered more attention. Best reported performance is 92% accuracy [93], but with a small dataset of only 100 subjects. In general, works that focus on this problem use low amounts of data, probably because such diagnostic groups are more sparse in available public and private cohorts, since a long follow-up is needed to correctly assess whether the patient will progress to AD.

We do not observe large differences in performance between modalities. Multimodal studies obtain strong results [34, 42, 98, 137], but they do not excel in any of the problems. Even though some papers show better performance by combining modalities [34], and others show good results [21, 41, 42, 137], they only achieve marginal improvements compared to single modality works. This could happen because 1) the models used do not completely leverage the data in all the modalities, or 2) true comparisons between methods are not reliable due to disparity in test/training sets. Models that include cognitive assessments in their analysis report strong results in detecting patients that will progress from MCI to AD [21, 40, 42, 56, 98, 101, 137].

It is difficult to compare classification performance, since each paper uses different data for testing and different approaches to validate its results. Overfitting is also a common problem [138, 139], where the methods presented perform well for a specific dataset but do not generalize to unseen data. For example, [57] performs the best in the MCI/CN classification task combining functional and structural MRI, but it uses a low amount of data and it does an exhaustive search of parameters, so the results are probably overfitted to the dataset. In our reviewed papers, the most used form of validation is cross-validation, and some papers also use other external datasets to test their model [84, 112]. When the amount of data is low, leave one out cross-validation is also an option. When possible, it is recommended to use such techniques for features election and hyperparameter optimization, as well as independent tests sets when comparing methods or model architectures.

Many studies suffer from reproducibility issues (Section 7, where experiments, methodologies and data are not detailed enough for other researchers to reproduce the same results. Some researchers have argued for test data standardization [138, 140, 141], so that the obtained results can be directly compared [31, 42, 137]. Sharing the code used for the experiments whenever possible and providing the information to reconstruct the exact dataset used, even if data itself are not public, should be a priority to tackle this problem.

6. Methodology challenges

Longitudinal data studies pose methodology challenges in both data collection and analysis (see Appendix C. In this section, we focus on two data analysis challenges: temporal alignment (i.e., baseline adjustment) and missing data; reporting how the reviewed works have addressed these issues and discussing some best practices to overcome them.

6.1. Temporal alignment

To consistently compare patients in a clinically meaningful way, the acquired markers should be temporally aligned. However, different patients can be at different stages of the disease at the same acquisition time, and data may not be necessarily acquired at the same biological age (i.e. degradation due to disease) for all subjects. Moreover, the gathered data only show a small snippet of the full onset of the disease (which can be up to 20 years). This is an important challenge that needs to be addressed, and it is especially critical in progression modelling and temporal prediction of the disease. Some of the reference variables used in the literature are time from baseline acquisition [47, 114], age of the subject [142], normalized age (where age variation has been removed [127]), cognitive scores [38, 116], data-driven progression scores [45, 82, 83, 87, 129] or any other type of data-driven temporal alignment [123].

Methods that account for alignment provide a more complete interpretation of the disease progression over time, as they allow studying between-subjects differences and assessing the evolution of a given patient with respect to that of the general population. Temporal alignment of patients could be tackled in two different ways: as a preprocessing step to a progression model or as a standalone problem. Development of new data-driven alignment methods is key to advance in this field. We believe that works such as [76, 123], or EBM related methods [28, 63, 110] show how useful data-driven approaches can be to this problem.

6.2. Missing data

Incomplete data are very common in longitudinal and/or multimodal clinical studies. Methods must deal with missing data to avoid possible biases, fully leverage available data, and be applicable to situations where some data could not be available. Depending on the pattern of missingness, there are different types of missing data in longitudinal studies [143]. It is generally assumed that data are missing at random, and most research has focused on strategies to handle this type of data.

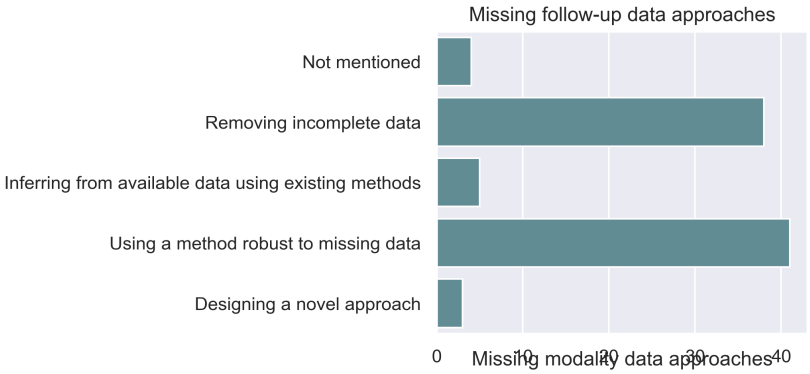

Figure 7: Distribution of papers by their approach to missing longitudinal data.

Figure 7 shows the distribution of the works according to the strategy used to deal with missing longitudinal data. We have divided the different approaches

in five categories:

- **Not mentioned:** Papers in this category did not report their approach to missing data. It is implied that they used a data cohort without missing entries. However, given the importance of this problem, it is concerning that it was not even mentioned.

- **Removing incomplete data:** A solution is selecting only a balanced subset of the available data, removing patients with missing longitudinal acquisitions. This approach has two main problems: it reduces the amount of available data, discarding potentially relevant information, and it can introduce biases in the data, especially if they are not missing at random [144].

- **Inferring from available data using existing methods:** Some works impute missing data from the available cohort, using simple methods such as average value [55] or direct completion from previous time points [134, 145], or more complex methods such as sparse regression [100] or low-rank matrix completion [53]. These approaches can be useful to deal with small amounts of random missing data, and they can be used as a preprocessing step, but tend to not scale well and become imprecise with larger amounts [14].

- **Using a method robust to missing data:** In these papers, the method itself accounts for unbalanced data [143]. For example, approaches based on mixed effect models, which are robust to missing observations [114, 129], or approaches where each time point is processed separately [65]. Some studies have adapted methods that were not initially flexible to unbalanced datasets. For example, in [145] they proposed a loss function for their model where only available data were used for its calculation, and [95] defined a temporally-structured SVM where different amount of

follow-ups could be used. This approach is more complex than just inferring the data, but can lead to more robust models and to the adaptation of existing methods.

- **Designing a novel approach:** Some works propose novel approaches to data missingness, making it a central point of their work [15, 102, 105, 122, 123]. Those that work with unbalanced data are more broadly applicable: a larger dataset can be used for training/validating the model, and it can be transferred more easily on a clinical setting, where the available data for a given subject may be sparse.

Although the majority of the reviewed papers address this important issue, a sizeable proportion (42.8%) just focus on analyzing curated datasets. This can be useful to showcase new methods and concepts, but not for creating a model that works in a real environment. Adapting existing (cross-sectional) models, such as SVM or deep learning, to the longitudinal domain [15, 95, 102, 145] could be a stepping stone to developing novel ML models.

Figure 8: Distribution of the reviewed papers (categorized by their main aplication) according to whether they deal with longitudinal missing data.

Figure 8 shows the proportion of reviewed papers that deal with missing data (inferring it or using a novel approach or a method robust to it) according to their main objective (either classification for CAD or modelling disease progression). Whereas methods for modelling disease progression can usually deal with unbalanced longitudinal data, classification methods for CAD are more sensitive to missing data.

Multimodal studies need to deal with missing data across modalities, as well as across time. Figure 9 shows the distribution of works according to the strategy used to handle missing data in multimodal studies. Some of them also appeared in Figure 7, as they deal with both types of data imbalance [39, 82, 123, 145]. Almost half of the studies chose not to use subjects with incomplete data, whereas the other half used methods that were more flexible. As before, progression models based on mixed effect models and other statistical modelling approaches are robust [39, 82, 113, 129], whereas supervised learning approaches for classification are more rigid [32, 99, 110].

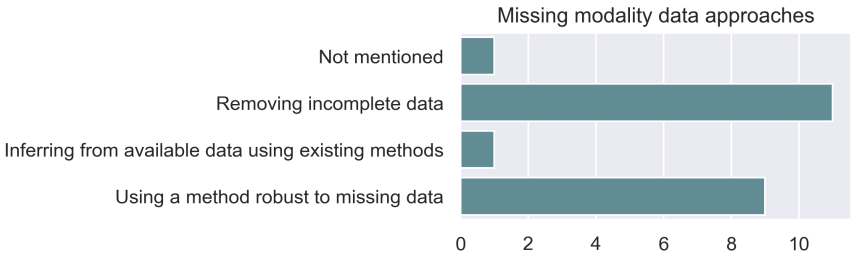

Figure 9: Distribution of papers by their approach to missing multimodal data. Only multimodal works were considered.

A considerable number of articles [16, 17, 29, 33, 46, 72, 118, 119] used simulated data to test their algorithm for other types of data. This approach allows researchers to create specific scenarios to evaluate the robustness of their algorithms; for example, with large amount of missing data, with only short-term longitudinal data, or with additional imaging modalities.

7. Reproducibility and interpretability

For all studies, results and methods should be reproducible to 1) ensure that the results are legitimate, 2) be able to directly apply the method to other datasets and 3) facilitate dissemination and open science. Moreover, if the model is to improve diagnosis or better understand the progression of the disease, it needs to be open and interpretable, that is, able to identify the factors that are responsible for triggering a concrete response.

Some measures to ensure the reproducibility of the published results and methods are:

- Using standardized datasets: some studies [31, 138, 141, 146] propose or use a concrete set of patients so that anyone can work with the same data. However, such standardized datasets have yet to be widely adopted, as the majority of the reviewed works either use private datasets, or do not precise which data are used from public datasets.

- Using a standardized data management for neuroimage storing and sharing, such as BIDS [1].

- Using methods such as cross-validation to minimize overfitting [57].

- Reproducing the methods on a completely different cohort of patients to test the robustness of the results [83].

- Making the code used to generate the code publicly available.

We studied the reproduciblity of the reviewed works by checking if the code and the data to generate the model and the reported results are available. We

[1] `https://bids.neuroimaging.io/`

considered that data were reported if they were directly available, or they could be obtained without ambiguity from a public dataset. As shown in Figure 10, although some papers make their data available, very few (15.3%) include the code. Most papers report neither data nor code.

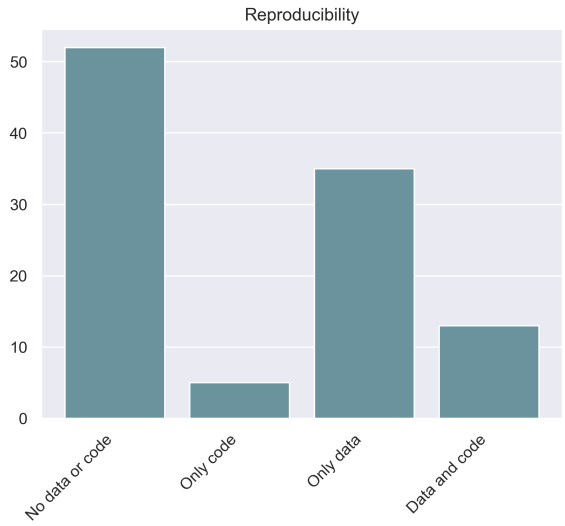

Figure 10: Distribution of the reviewed papers by their reproducibility. Vertical axis indicates amount of papers.

Regarding interpretability, some ML methods, such as deep learning, have been criticized for being a black box [147], not providing the underlying reasoning of the model. Interestingly, many of the reviewed works chose ML methods allowing some level of interpretability, which is crucial for clinical use. Most of the disease progression models are interpretable since their underpinning aim is to obtain a clinically interpretable representation of the disease pathophysiology. Other examples are methods based on MKL [98, 99, 136], where markers are assigned a weight according to their importance to the learned task, or manifold learning methods [65–67], where patients can be embedded in a low-dimensional space to directly visualize relationships between them.

8. Discussion

We have surveyed papers that use ML algorithms for longitudinal data analysis. Most of the works focus on neuroimaging, mainly MRI. A significant percentage of the works (approximately 26%) use multimodal data. Although reviewed papers target various tasks, we can divide them in two groups: disease progression modelling and CAD. Methods for disease progression modelling can usually deal with unbalanced longitudinal data, whereas methods for CAD are more sensitive to missing data.

For some specific problems, such as classification between converting and non-converting MCI, longitudinal data showed a strong performance [52, 93], compared to standard cross-sectional methods [138]. Better detection of MCI converters could led to an improved early detection of the disease, through development of novel methods that help prevention policies, and longitudinal epidemiological studies that focus on early stages and healthy subjects. Since AD affects different biological processes, multimodal studies provide a more comprehensive view of the disease. Studies using more than one modality are gaining importance [34, 41, 42, 98, 99], and results in specific problems, such as classification in CAD systems, have shown a slight improvement with respect to single modality studies. There is enough evidence from existing multimodal studies [39, 64, 82] that integrating those different sources can boost performance.

Based on our analysis, we argue that methods should aim for robustness to missing and unbalanced data, especially for CAD applications. Among the different approaches to tackle this problem, a promising one is using simulated data [16, 17, 29, 33, 46, 72, 118, 119]. This allows validating a model in different settings and testing its robustness for different rates of unbalanced/missing data, or setting baseline before using real data.

To improve understanding of the disease and make the methods rigorous and applicable to a clinical setting, more effort towards reproducibility and interpretability of the methods and results is needed. A more widespread use of validation tools and cohorts would be desirable. Many papers do not use standardized datasets, nor share their data, so their results are hard to compare against other works. Moreover, some of the papers rely on "hard-to-interpret" ML techniques [54, 59].

Despite large advances on longitudinal data analysis, more research is necessary on methods that better process and interpret the huge influx of relevant information that a longitudinal characterization of the disease can offer. Prior knowledge or assumptions of the disease should be incorporated to naturally accommodate longitudinal data [63, 89, 95]. In this context, the ATN biomarker framework [148], which provides a biologically based definition of AD (see Appendix B, opens up a new path to more accurately characterize the disease.

Regarding CAD applications, deep learning approaches have achieved great success in medical imaging [108], in brain disease diagnosis [149, 150] and more specifically, in AD [151, 152], using cross-sectional imaging data. However, our findings show that their application for longitudinal analysis is still low (only 6% of the reviewed articles used deep learning based techniques). Incorporate longitudinal neuroimaging data to deep learning is challenging to classification systems, due to the high dimensionality that a temporal dimension adds. Moreover, it is not straightforward to solve the problem of missing data and variable number of follow-ups in a multi-layer architecture, as several works addressing this problem show [15, 59]. More work should be done to incorporate such techniques to the study of longitudinal, high dimensional neuroimaging data, where they hold promise for better understanding and treatment of the disease. Given the aforementioned success and great performance of deep learning for cross-sectional studies, we encourage and expect advancements in deep learning based systems using longitudinal data in the near future.

Almost all reviewed papers use supervised methods. Unsupervised (or semi-supervised) learning could be useful in dealing with longitudinal data, where data are usually unbalanced or not labelled in some of the modalities. Clustering or other unsupervised techniques, for example, could be used to study the relationships between different trajectories of patients, without relying on labels or on large amounts of data.

More research to overcome the described problems and challenges is key to broaden our understanding about the progression of AD and other neurodegenerative diseases. These methodological advances would open the door to develop applications that can be useful in clinical and epidemiological settings.

Acknowledgements

This research was partially funded by the "Fundació La Marató de TV3" (nº20154031). This work was also funded by the Spanish Ministry of Economy and Competitiveness under the María de Maeztu Units of Excellence Programme [MDM-2015-0502]. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C. P. Ferri, M. Prince, C. Brayne, H. Brodaty, L. Fratiglioni, M. Ganguli, K. Hall, K. Hasegawa, H. Hendrie, Y. Huang, A. Jorm, C. Mathers, P. R. Menezes, E. Rimmer, M. Scazufca, Alzheimer's Disease International, Global prevalence of dementia: a Delphi consensus study, Lancet 366 (2005) 2112–2117.

[2] M. Prince, R. Bryce, E. Albanese, A. Wimo, W. Ribeiro, C. P. Ferri, The global prevalence of dementia: A systematic review and metaanalysis, Alzheimer's Dement. 9 (2013) 63–75.e2.

[3] C. Ballard, S. Gauthier, A. Corbett, C. Brayne, D. Aarsland, E. Jones, Alzheimer's disease, Lancet 377 (2011) 1019–1031.

[4] L. Bäckman, S. Jones, A. K. Berger, E. J. Laukka, B. J. Small, Multiple cognitive deficits during the transition to Alzheimer's disease, J. Intern. Med. 256 (2004) 195–204.

[5] C. R. Jack, D. S. Knopman, W. J. Jagust, L. M. Shaw, P. S. Aisen, M. W. Weiner, R. C. Petersen, J. Q. Trojanowski, Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade, Lancet Neurol. 9 (2010) 119–128.

[6] E. Lawrence, C. Vegvari, A. Ower, C. Hadjichrysanthou, F. De Wolf, R. M. Anderson, A Systematic Review of Longitudinal Studies Which Measure Alzheimer's Disease Biomarkers, J. Alzheimer's Dis. 59 (2017) 1359–1379.

[7] K. L. Mills, C. K. Tamnes, Methods and considerations for longitudinal structural brain imaging analysis across development, Dev. Cogn. Neurosci. 9 (2014) 172–190.

[8] L. Breiman, Statistical Modeling: The Two Cultures, Stat. Sci. 16 (2001) 199–231.

[9] G. Shmueli, To Explain or to Predict?, Stat. Sci. 25 (2010) 289–310.

[10] N. P. Oxtoby, D. C. Alexander, Imaging plus X: Multimodal models of neurodegenerative disease, Curr. Opin. Neurol. 30 (2017) 371–379.

[11] U. R. Acharya, S. L. Fernandes, J. E. WeiKoh, E. J. Ciaccio, M. K. M. Fabell, U. J. Tanik, V. Rajinikanth, C. H. Yeong, Automated Detection

of Alzheimer's Disease Using Brain MRI Images– A Study with Various Feature Extraction Techniques, J. Med. Syst. 43 (2019).

[12] Z. Xu, X. Shen, W. Pan, Longitudinal analysis is more powerful than cross-sectional analysis in detecting genetic association with neuroimaging phenotypes, PLoS One 9 (2014).

[13] G. M. Fitzmaurice, C. Ravichandran, A primer in longitudinal data analysis, Circulation 118 (2008) 2005–2010.

[14] J. G. Ibrahim, G. Molenberghs, Missing data methods in longitudinal studies: a review., Test (Madr). 18 (2009) 1–43.

[15] M. Mehdipour Ghazi, M. Nielsen, A. Pai, M. J. Cardoso, M. Modat, S. Ourselin, L. Sørensen, Training recurrent neural networks robust to incomplete data: Application to Alzheimer's disease progression modeling, Med. Image Anal. 53 (2019) 39–46.

[16] R. Chen, S. M. Resnick, C. Davatzikos, E. H. Herskovits, Dynamic Bayesian network modeling for longitudinal brain morphometry., Neuroimage 59 (2012) 2330–2338.

[17] K. Li, S. Luo, Functional joint model for longitudinal and time-to-event data: an application to Alzheimer's disease, Stat. Med. 36 (2017) 3560–3572.

[18] K. Li, W. Chan, R. S. Doody, J. Quinn, S. Luo, Prediction of Conversion to Alzheimer's Disease with Longitudinal Measures and Time-To-Event Data, J. Alzheimer's Dis. 58 (2017) 361–371.

[19] M. Bilgel, B. Jedynak, D. F. Wong, S. M. Resnick, J. L. Prince, Temporal trajectory and progression score estimation from voxelwise longitudinal imaging measures: Application to amyloid imaging, in: Lect. Notes Comput. Sci., volume 9123, NIH Public Access, 2015, pp. 424–436. doi:10.1007/978-3-319-19992-4_33.

[20] M. Bilgel, J. L. Prince, D. F. Wong, S. M. Resnick, B. M. Jedynak, A multivariate nonlinear mixed effects model for longitudinal image analysis: Application to amyloid imaging., Neuroimage 134 (2016) 658–670.

[21] B. A. Ardekani, E. Bermudez, A. M. Mubeen, A. H. Bachman, Prediction of incipient Alzheimer's disease dementia in patients with mild cognitive impairment, J. Alzheimer's Dis. 55 (2016) 269–281.

[22] J. B. Fiot, L. Risser, L. D. Cohen, J. Fripp, F. X. Vialard, Local vs global descriptors of hippocampus shape evolution for Alzheimer's longitudinal population analysis, Lect. Notes Comput. Sci. 7570 (2012) 13–24.

[23] J.-B. Fiot, H. Raguet, L. Risser, L. D. Cohen, J. Fripp, F.-X. Vialard, Longitudinal deformation models, spatial regularizations and learning strategies to quantify Alzheimer's disease progression., NeuroImage. Clin. 4 (2014) 718–729.

[24] K. R. Gray, R. Wolz, R. A. Heckemann, P. Aljabar, A. Hammers, D. Rueckert, Multi-region analysis of longitudinal FDG-PET for the classification of Alzheimer's disease, Neuroimage 60 (2012) 221–229.

[25] F. Rodrigues, M. Silveira, Longitudinal FDG-PET features for the classification of Alzheimer's disease., Conf. Proc. ... Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf. 2014 (2014) 1941–1944.

[26] B. Shi, Y. Chen, C. Science, K. Hobbs, C. D. Smith, J. Liu, Nonlinear Metric Learning for Alzheimer ' s Disease Diagnosis with Integration of Longitudinal Neuroimaging Features, BMVC2015 (2015) 1–13.

[27] B. Shi, Y. Chen, P. Zhang, C. D. Smith, J. Liu, Nonlinear feature transformation and deep fusion for Alzheimer's Disease staging analysis, Pattern Recognit. 63 (2017) 487–498.

[28] J. Huang, D. Alexander, Probabilistic Event Cascades for Alzheimer's disease, Adv. Neural Inf. Process. Syst. 25 (2012) 3104–3112.

[29] J. W. Hyun, Y. Li, C. Huang, M. Styner, W. Lin, H. Zhu, STGP: Spatio-temporal Gaussian process models for longitudinal neuroimaging data, Neuroimage 134 (2016) 550–562.

[30] L. K. McEvoy, D. Holland, D. J. Hagler, C. Fennema-Notestine, J. B. Brewer, A. M. Dale, Mild cognitive impairment: baseline and longitudinal structural MR imaging measures improve predictive prognosis., Radiology 259 (2011) 834–843.

[31] G. Sanroma, V. Andrea, O. M. Benkarim, J. V. Manjón, P. Coupé, O. Camara, G. Piella, M. A. González Ballester, Early prediction of alzheimer's disease with non-local patch-based longitudinal descriptors, in: Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), volume 10530 LNCS, Springer, Cham, 2017, pp. 74–81. doi:10.1007/978-3-319-67434-6_9.

[32] M. Vounou, E. Janousova, R. Wolz, J. L. Stein, P. M. Thompson, D. Rueckert, G. Montana, Sparse reduced-rank regression detects genetic associations with voxel-wise longitudinal phenotypes in Alzheimer's disease., Neuroimage 60 (2012) 700–716.

[33] S. Chen, F. Dubois Bowman, A novel support vector classifier for longitudinal high-dimensional data and its application to neuroimaging data, Stat. Anal. Data Min. 4 (2011) 604–611.

[34] C. Hinrichs, V. Singh, G. Xu, S. C. Johnson, Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population, Neuroimage 55 (2011) 574–589.

[35] C. R. Jack, V. J. Lowe, S. D. Weigand, H. J. Wiste, M. L. Senjem, D. S. Knopman, M. M. Shiung, J. L. Gunter, B. F. Boeve, B. J. Kemp, M. Weiner, R. C. Petersen, Serial PIB and MRI in normal, mild cognitive impairment and Alzheimers disease: Implications for sequence of pathological events in Alzheimers disease, Brain 132 (2009) 1355–1365.

[36] R. S. Desikan, L. K. McEvoy, W. K. Thompson, D. Holland, J. C. Rddey, K. Blennow, P. S. Aisen, J. B. Brewer, B. T. Hyman, A. M. Dale, Amyloid-β associated volume loss occurs only in the presence of phospho-tau, Ann. Neurol. 70 (2011) 657–661.

[37] B. Guillaume, X. Hua, P. M. Thompson, L. Waldorp, T. E. Nichols, Fast and accurate modelling of longitudinal and repeated measures neuroimaging data, Neuroimage 94 (2014) 287–302.

[38] R. Guerrero, A. Schmidt-Richberg, C. Ledig, T. Tong, R. Wolz, D. Rueckert, Instantiated mixed effects modeling of Alzheimer's disease markers, Neuroimage 142 (2016) 113–125.

[39] Y. Iturria-Medina, R. C. Sotero, P. J. Toussaint, J. M. Mateos-Pérez, A. C. Evans, Early role of vascular dysregulation on late-onset Alzheimer's disease based on multifactorial data-driven analysis, Nat. Commun. 7 (2016) 11934.

[40] M. Aghili, S. Tabarestani, M. Adjouadi, E. Adeli, Predictive Modeling of Longitudinal Data for Alzheimer's Disease Diagnosis Using RNNs, in: Predict. Intell. Med., Springer International Publishing, 2018, pp. 112–119. doi:10.1007/978-3-030-00320-3_14.

[41] C. L. Chi, W. Zeng, W. Oh, S. Borson, T. Lenskaia, X. Shen, P. J. Tonellato, Personalized long-term prediction of cognitive function: Using sequential assessments to improve model performance, J. Biomed. Inform. 76 (2017) 78–86.

[42] S. Minhas, A. Khanum, F. Riaz, A. Alvi, S. S. A. Khan, Minhas, Khanum, Riaz, A. Alvi, Khan, S. Minhas, A. Khanum, F. Riaz, A. Alvi, S. S. A. Khan, A Non Parametric Approach for Mild Cognitive Impairment to AD Conversion Prediction: Results on Longitudinal Data., IEEE J. Biomed. Heal. Informatics 21 (2016) 1–2.

[43] A. L. Young, N. P. Oxtoby, J. Huang, R. V. Marinescu, P. Daga, D. M. Cash, N. C. Fox, S. Ourselin, J. M. Schott, D. C. Alexander, Multiple orderings of events in disease progression, in: Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), volume 9123, Springer, Cham, 2015, pp. 711–722. doi:10.1007/978-3-319-19992-4_56.

[44] M. W. Weiner, D. P. Veitch, P. S. Aisen, L. A. Beckett, N. J. Cairns, R. C. Green, D. Harvey, C. R. Jack, W. Jagust, J. C. Morris, R. C. Petersen, A. J. Saykin, L. M. Shaw, A. W. Toga, J. Q. Trojanowski, Recent publications from the Alzheimer's Disease Neuroimaging Initiative: Reviewing progress toward improved AD clinical trials, 2017. doi:10.1016/j.jalz.2016.11.007.

[45] C. Davatzikos, F. Xu, Y. An, Y. Fan, S. M. Resnick, Longitudinal progression of Alzheimers-like patterns of atrophy in normal older adults: The SPARE-AD index, Brain 132 (2009) 2026–2035.

[46] R. V. R. V. Marinescu, A. Eshaghi, M. Lorenzi, A. L. Young, N. P. Oxtoby, S. Garbarino, T. J. Shakespeare, S. J. Crutch, D. C. Alexander, R. V. R. V. Marinescu, A. Eshaghi, M. Lorenzi, A. L. Young, N. P. Oxtoby, S. Garbarino, T. J. Shakespeare, S. J. Crutch, D. C. Alexander, A Vertex Clustering Model for Disease Progression: Application to Cortical Thickness Images, in: M. Niethammer, M. Styner, S. Aylward, H. Zhu, I. Oguz, P.-T. Yap, D. Shen (Eds.), Inf. Process. Med. Imaging, volume 10265, Springer International Publishing, Cham, 2017, pp. 134–145. doi:10.1007/978-3-319-59050-9.

[47] K. Franke, C. Gaser, Longitudinal Changes in Individual BrainAGE in Healthy Aging, Mild Cognitive Impairment, and Alzheimer's Disease, GeroPsych (Bern). 25 (2012) 235–245.

[48] M. A. Ikram, G. G. Brusselle, S. D. Murad, C. M. van Duijn, O. H. Franco, A. Goedegebure, C. C. Klaver, T. E. Nijsten, R. P. Peeters, B. H. Stricker,

H. Tiemeier, A. G. Uitterlinden, M. W. Vernooij, A. Hofman, The Rotterdam Study: 2018 update on objectives, design and main results, Eur. J. Epidemiol. 32 (2017) 807–850.

[49] N. Andreasen, L. Minthon, E. Vanmechelen, H. Vanderstichele, P. Davidsson, B. Winblad, K. Blennow, Cerebrospinal fluid tau and Aβ42 as predictors of development of Alzheimer's disease in patients with mild cognitive impairment, Neurosci. Lett. 273 (1999) 5–8.

[50] C. M. Clark, J. A. Schneider, B. J. Bedell, T. G. Beach, W. B. Bilker, M. A. Mintun, M. J. Pontecorvo, F. Hefti, A. P. Carpenter, M. L. Flitter, M. J. Krautkramer, H. F. Kung, R. E. Coleman, P. M. Doraiswamy, A. S. Fleisher, M. N. Sabbagh, C. H. Sadowsky, P. E. M. Reiman, S. P. Zehntner, D. M. Skovronsky, Use of florbetapir-PET for imaging β-amyloid pathology, J. Am. Med. Assoc. 305 (2011) 275–283.

[51] M. Weiner, Z. Khachaturian, The Use of MRI and PET for Clinical Diagnosis of Dementia and Investigation of Cognitive Impairment: A Consensus Report, Alzheimer's Assoc. Chicago, (2005) 1–15.

[52] M. Liu, H. I. Suk, D. Shen, Multi-task sparse classifier for diagnosis of MCI conversion to AD with longitudinal MR images, in: Mach. Learn. Med. Imaging (MLMI). Lect. Notes Comput. Sci., volume 8184, 2013, pp. 243–250. doi:10.1007/978-3-319-02267-3_31.

[53] K.-H. Thung, C.-Y. Wee, P.-T. Yap, D. Shen, Identification of progressive mild cognitive impairment patients using incomplete longitudinal MRI scans., Brain Struct. Funct. 221 (2016) 3979–3995.

[54] J. Zhang, Q. Li, R. J. Caselli, J. Ye, Y. Wang, Multi-task Dictionary Learning based Convolutional Neural Network for Computer aided Diagnosis with Longitudinal Images, ArXiv (2017).

[55] J. Zhou, J. Liu, V. A. Narayan, J. Ye, Modeling disease progression via multi-task learning, Neuroimage 78 (2013) 233–248.

[56] K.-H. Thung, P.-T. Yap, E. Adeli, S.-W. Lee, D. Shen, Conversion and time-to-conversion predictions of mild cognitive impairment using low-rank affinity pursuit denoising and matrix completion, Med. Image Anal. 45 (2018) 68–82.

[57] P. Yang, D. Ni, S. Chen, T. Wang, D. Wu, B. Lei, Multi-task fused sparse learning for mild cognitive impairment identification., Technol. Health Care 26 (2018) 437–448.

[58] L. M. Aksman, M. A. Scelsi, A. F. Marquand, D. C. Alexander, S. Ourselin, A. Altmann, Modeling longitudinal imaging biomarkers with parametric Bayesian multi-task learning, Hum. Brain Mapp. (2019) hbm.24682.

[59] L. E. Givon, L. J. Mariano, D. O'Dowd, J. M. Irvine, A. R. Schneider, Cognitive Subscore Trajectory Prediction in Alzheimer's Disease, ArXiv (2017).

[60] A. Ortiz, J. Munilla, F. J. Martínez-Murcia, J. M. Górriz, J. Ramírez, Learning longitudinal MRI patterns by SICE and deep learning: Assessing the Alzheimer's disease progression, in: Commun. Comput. Inf. Sci., volume 723, Springer, Cham, 2017, pp. 413–424. doi:10.1007/978-3-319-60964-5_36.

[61] R. Cui, M. Liu, G. Li, Longitudinal analysis for Alzheimer's disease diagnosis using RNN, Proc. - Int. Symp. Biomed. Imaging 2018-April (2018) 1398–1401.

[62] N. Bhagwat, J. D. Viviano, A. N. Voineskos, M. M. Chakravarty, Modeling and prediction of clinical symptom trajectories in Alzheimer's disease using longitudinal data, PLoS Comput. Biol. 14 (2018) e1006376.

[63] H. M. Fonteijn, M. Modat, M. J. Clarkson, J. Barnes, M. Lehmann, N. Z. Hobbs, R. I. Scahill, S. J. Tabrizi, S. Ourselin, N. C. Fox, D. C. Alexander,

An event-based model for disease progression and its application in familial Alzheimer's disease and Huntington's disease, Neuroimage 60 (2012) 1880–1889.

[64] N. P. Oxtoby, A. L. Young, D. M. Cash, T. L. S. Benzinger, A. M. Fagan, J. C. Morris, R. J. Bateman, N. C. Fox, J. M. Schott, D. C. Alexander, Data-driven models of dominantly-inherited Alzheimer's disease progression, Brain A J. Neurol. 141 (2018) 1529 – 1544.

[65] R. Guerrero, C. Ledig, A. Schmidt-Richberg, D. Rueckert, Group-constrained manifold learning: Application to AD risk assessment, Pattern Recognit. 63 (2017) 570–582.

[66] R. Wolz, P. Aljabar, J. V. Hajnal, D. Rueckert, Manifold learning for biomarker discovery in MR imaging, Mach. Learn. Med. Imaging 6357 (2010) 116–123.

[67] R. Guerrero, C. Ledig, A. Schmidt-Richberg, D. Rueckert, Group-constrained Laplacian Eigenmaps: Longitudinal AD biomarker learning, in: Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), volume 9352, Springer International Publishing, 2015, pp. 178–185. doi:10.1007/978-3-319-24888-2_22.

[68] H. H. Dodge, J. Zhu, D. Harvey, N. Saito, L. C. Silbert, J. A. Kaye, R. A. Koeppe, R. L. Albin, Biomarker progressions explain higher variability in stage-specific cognitive decline than baseline values in Alzheimer disease, Alzheimer's Dement. 10 (2014) 690–703.

[69] I. Koval, J.-B. Schiratti, A. Routier, M. Bacci, O. Colliot, S. Allassonnière, S. Durrleman, Spatiotemporal Propagation of the Cortical Atrophy: Population and Individual Patterns, Front. Neurol. 9 (2018) 235.

[70] D. M. Cash, C. Frost, L. O. Iheme, D. Ünay, M. Kandemir, J. Fripp, O. Salvado, P. Bourgeat, M. Reuter, B. Fischl, M. Lorenzi, G. B. Frisoni, X. Pennec, R. K. Peirson, J. L. Gunter, M. L. Senjem, C. R. Jack,

N. Guizard, V. S. Fonov, D. L. Collins, M. Modat, M. J. Cardoso, K. K. Leung, H. Wang, S. R. Das, P. A. Yushkevich, I. B. Malone, N. C. Fox, J. M. Schott, S. Ourselin, Assessing atrophy measurement techniques in dementia: Results from the MIRIAD atrophy challenge, Neuroimage 123 (2015) 149–164.

[71] J. B. Schiratti, S. Allassonniere, A. Routier, S. Durrleman, A Mixed-Effects Model with Time Reparametrization for Longitudinal Univariate Manifold-Valued Data., Inf. Process. Med. Imaging 24 (2015) 564–575.

[72] G. Ziegler, W. D. Penny, G. R. Ridgway, S. Ourselin, K. J. Friston, Estimating anatomical trajectories with Bayesian mixed-effects modeling, Neuroimage 121 (2015) 51–68.

[73] L. Gui, X. Tang, J. M. F. Moura, Geodesic distance on a Grassmannian for monitoring the progression of Alzheimer's disease., Neuroimage 146 (2017) 1016–1024.

[74] X. Tang, D. Holland, A. M. Dale, L. Younes, M. I. Miller, The diffeomorphometry of regional shape change rates and its relevance to cognitive deterioration in mild cognitive impairment and Alzheimer's disease, Hum. Brain Mapp. 36 (2015) 2093–2117.

[75] A. Bône, M. Louis, A. Routier, J. Samper, M. Bacci, B. Charlier, O. Colliot, S. Durrleman, Prediction of the progression of subcortical brain structures in Alzheimer's disease from baseline, Lect. Notes Comput. Sci. 10551 (2017) 101–113.

[76] A. Bône, O. Colliot, S. Durrleman, Learning distributions of shape trajectories from longitudinal datasets: a hierarchical model on a manifold of diffeomorphisms, in: CVPR 2018 - Comput. Vis. Pattern Recognit. 2018, Salt Lake City, United States, 2018, pp. 9271–9280. arXiv:1803.10119.

[77] B. A. Gutman, X. Hua, P. Rajagopalan, Y.-Y. Chou, Y. Wang, I. Yanovsky, A. W. Toga, C. R. J. Jack, M. W. Weiner, P. M. Thompson,

Maximizing power to track Alzheimer's disease and MCI progression by LDA-based weighting of longitudinal ventricular surface features., Neuroimage 70 (2013) 386–401.

[78] M. Lorenzi, G. Ziegler, D. C. Alexander, S. Ourselin, Efficient Gaussian Process-Based Modelling and Prediction of Image Time Series, Inf. Process. Med. Imaging 24 (2015) 626–637.

[79] M. Lorenzi, G. Ziegler, D. C. Alexander, S. Ourselin, Modelling non-stationary and non-separable spatio-temporal changes in neurodegeneration via gaussian process convolution, Lect. Notes Comput. Sci. 9487 (2015) 35–44.

[80] M. Lorenzi, M. Filippone, G. B. Frisoni, D. C. Alexander, S. Ourselin, Disease Progression Modeling and Prediction through Random Effect Gaussian Processes and Time Transformation, Neuroimage (2017).

[81] C. Gaser, K. Franke, S. Klöppel, N. Koutsouleris, H. Sauer, BrainAGE in Mild Cognitive Impaired Patients: Predicting the Conversion to Alzheimer's Disease, PLoS One 8 (2013).

[82] B. M. Jedynak, A. Lang, B. Liu, E. Katz, Y. Zhang, B. T. Wyman, D. Raunig, P. Jedynak, B. Caffo, J. L. Prince, C. P. Jedynak, B. Caffo, J. L. Prince, A computational neurodegenerative disease progression score: Method and results with the Alzheimer's disease neuroimaging initiative cohort, Neuroimage 63 (2012) 1478–1486.

[83] R. Casanova, R. T. Barnard, S. A. Gaussoin, S. Saldana, K. M. Hayden, J. E. Manson, R. B. Wallace, S. R. Rapp, S. M. Resnick, M. A. Espeland, J.-C. Chen, Using high-dimensional machine learning methods to estimate an anatomical risk factor for Alzheimer's disease across imaging databases., Neuroimage 183 (2018) 401–411.

[84] L. M. Aksman, D. J. Lythgoe, S. C. R. Williams, M. Jokisch, C. Monninghoff, J. Streffer, K.-H. Jockel, C. Weimar, A. F. Marquand, Making

use of longitudinal information in pattern recognition., Hum. Brain Mapp. 37 (2016) 4385–4404.

[85] C. Cabral, P. M. Morgado, D. Campos Costa, M. Silveira, Predicting conversion from MCI to AD with FDG-PET brain images at different prodromal stages, Comput. Biol. Med. 58 (2015) 101–109.

[86] A. Chincarini, F. Sensi, L. Rei, G. Gemme, S. Squarcia, R. Longo, F. Brun, S. Tangaro, R. Bellotti, N. Amoroso, M. Bocchetta, A. Redolfi, P. Bosco, M. Boccardi, G. B. Frisoni, F. Nobili, Integrating longitudinal information in hippocampal volume measurements for the early detection of Alzheimer's disease, Neuroimage 125 (2016) 834–847.

[87] V. H. Clark, S. M. Resnick, J. Doshi, L. L. Beason-Held, Y. Zhou, L. Ferrucci, D. F. Wong, M. A. Kraut, C. Davatzikos, Longitudinal imaging pattern analysis (SPARE-CD index) detects early structural and functional changes before cognitive decline in healthy older adults., Neurobiol. Aging 33 (2012) 2733–2745.

[88] A. Farzan, S. Mashohor, A. R. Ramli, R. Mahmud, Boosting diagnosis accuracy of Alzheimer's disease using high dimensional recognition of longitudinal brain atrophy patterns, Behav. Brain Res. 290 (2015) 124–130.

[89] G. Gavidia-Bovadilla, S. Kanaan-Izquierdo, M. Mataró-Serrat, A. Perera-Lluna, Early Prediction of Alzheimer's Disease Using Null Longitudinal Model-Based Classifiers, PLoS One 12 (2017) e0168011.

[90] B. Lei, F. Jiang, S. Chen, D. Ni, T. Wang, Longitudinal analysis for disease progression via simultaneous multi-relational temporal-fused learning, Front. Aging Neurosci. 9 (2017) 6.

[91] Y. Li, Y. Wang, G. Wu, F. Shi, L. Zhou, W. Lin, D. Shen, Discriminant analysis of longitudinal cortical thickness changes in Alzheimer's disease using dynamic and network features, Neurobiol. Aging 33 (2012) 15–30.

[92] C. Misra, Y. Fan, C. Davatzikos, Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: Results from ADNI, Neuroimage 44 (2009) 1415–1422.

[93] Z. Sun, M. van de Giessen, B. P. F. Lelieveldt, M. Staring, Detection of Conversion from Mild Cognitive Impairment to Alzheimer's Disease Using Longitudinal Brain MRI., Front. Neuroinform. 11 (2017) 16.

[94] J. Zhang, M. Liu, L. An, Y. Gao, D. Shen, Alzheimer's Disease Diagnosis Using Landmark-Based Features From Longitudinal Structural MR Images., IEEE J. Biomed. Heal. informatics 21 (2017) 1607–1616.

[95] Y. Zhu, X. Zhu, M. Kim, Early Diagnosis of Alzheimer's Disease by Joint Feature Selection and Classification on Temporally Structured Support Vector Machine, MICCAI 9900 (2016) 264–272.

[96] S. Adaszewski, J. Dukart, F. Kherif, R. Frackowiak, B. Draganski, How early can we predict Alzheimer's disease using computational anatomy?, Neurobiol. Aging 34 (2013) 2815–2826.

[97] H. Guan, T. Liu, J. Jiang, D. Tao, J. Zhang, H. Niu, W. Zhu, Y. Wang, J. Cheng, N. A. Kochan, H. Brodaty, P. Sachdev, W. Wen, Classifying MCI Subtypes in Community-Dwelling Elderly Using Cross-Sectional and Longitudinal MRI-Based Biomarkers., Front. Aging Neurosci. 9 (2017) 309.

[98] T. Chen, D. Zeng, Y. Wang, Multiple kernel learning with random effects for predicting longitudinal outcomes and data integration, Biometrics 71 (2015) 918–928.

[99] D. Zhang, D. Shen, Predicting future clinical changes of MCI patients using longitudinal and multimodal biomarkers., PLoS One 7 (2012) e33182.

[100] L. Huang, Y. Jin, Y. Gao, K. H. Thung, D. Shen, Longitudinal clinical score prediction in Alzheimer's disease with soft-split sparse regression based random forest, Neurobiol. Aging 46 (2016) 180–191.

[101] A. M. Mubeen, A. Asaei, A. H. Bachman, J. J. Sidtis, B. A. Ardekani, A six-month longitudinal evaluation significantly improves accuracy of predicting incipient Alzheimer's disease in mild cognitive impairment., J. Neuroradiol. 44 (2017) 381–387.

[102] B. Jie, M. Liu, J. Liu, D. Zhang, D. Shen, Temporally Constrained Group Sparse Learning for Longitudinal Data Analysis in Alzheimer's Disease., IEEE Trans. Biomed. Eng. 64 (2017) 238–249.

[103] H. Wang, F. Nie, H. Huang, J. Yan, S. Kim, High-Order Multi-Task Feature Learning to Identify Longitudinal Phenotypic Markers for Alzheimer's Disease Progression Prediction., Nips (2012) 1–9.

[104] X. Wang, D. Shen, H. Huang, Prediction of Memory Impairment with MRI Data: A Longitudinal Study of Alzheimer's Disease, Lect. Notes Comput. Sci. 9900 (2016) 273–281.

[105] Y. Zhu, M. R. Sabuncu, A Probabilistic Disease Progression Model for Predicting Future Clinical Outcome, ArXiV (2018).

[106] M. A. Araque Caballero, S. Kloppel, M. Dichgans, M. Ewers, Spatial Patterns of Longitudinal Gray Matter Change as Predictors of Concurrent Cognitive Decline in Amyloid Positive Healthy Subjects., J. Alzheimers. Dis. 55 (2017) 343–358.

[107] M. Yang, P. Yang, A. Elazab, W. Hou, X. Li, T. Wang, W. Zou, B. Lei, Join and Deep Ensemble Regression of Clinical Scores for Alzheimer's Disease Using Longitudinal and Incomplete Data, Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS 2018-July (2018) 1254–1257.

[108] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, C. I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.

[109] T. Wang, R. G. Qiu, M. Yu, Predictive Modeling of the Progression of Alzheimer's Disease with Recurrent Neural Networks, Sci. Rep. 8 (2018) 9161.

[110] A. L. Young, N. P. Oxtoby, P. Daga, D. M. Cash, N. C. Fox, S. Ourselin, J. M. Schott, D. C. Alexander, A data-driven model of biomarker changes in sporadic Alzheimer's disease., Brain 137 (2014) 2564–2577.

[111] J. L. Bernal-Rusiel, D. N. Greve, M. Reuter, B. Fischl, M. R. Sabuncu, Statistical analysis of longitudinal neuroimage data with Linear Mixed Effects models, Neuroimage 66 (2013) 249–260.

[112] C. Platero, L. Lin, M. C. Tobar, Longitudinal Neuroimaging Hippocampal Markers for Diagnosing Alzheimer's Disease, Neuroinformatics 17 (2019) 43–61.

[113] V. L. Villemagne, S. Burnham, P. Bourgeat, B. Brown, K. A. Ellis, O. Salvado, C. Szoeke, S. L. Macaulay, R. Martins, P. Maruff, D. Ames, C. C. Rowe, C. L. Masters, Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: A prospective cohort study, Lancet Neurol. 12 (2013) 357–367.

[114] M. C. Donohue, H. Jacqmin-Gadda, M. Le Goff, R. G. Thomas, R. Raman, A. C. Gamst, L. A. Beckett, C. R. Jack, M. W. Weiner, J. F. Dartigues, P. S. Aisen, Estimating long-term multivariate progression from short-term data, 2014. doi:10.1016/j.jalz.2013.10.003. arXiv:NIHMS150003.

[115] D. Li, S. Iddi, W. K. Thompson, M. C. Donohue, Bayesian latent time joint mixed effect models for multicohort longitudinal data, Stat. Methods Med. Res. (2017) 1–14.

[116] E. Yang, M. Farnum, V. Lobanov, T. Schultz, N. Raghavan, M. N. Samtani, G. Novak, V. Narayan, A. Dibernardo, Quantifying the pathophys-

iological timeline of Alzheimer's disease, J. Alzheimer's Dis. 26 (2011) 745–753.

[117] K.-Y. Liang, S. L. Zeger, Longitudinal data analysis using generalized linear models, Biometrika 73 (1986) 13–22.

[118] X. Zhang, L. Li, H. Zhou, D. Shen, t. A. D. N. Initiative, Tensor Generalized Estimating Equations for Longitudinal Imaging Analysis (2014).

[119] Y. Li, J. H. Gilmore, D. Shen, M. Styner, W. Lin, H. Zhu, Multiscale adaptive generalized estimating equations for longitudinal neuroimaging data, Neuroimage 72 (2013) 91–105.

[120] C. Wachinger, A. Rieckmann, M. Reuter, Latent processes governing neuroanatomical change in aging and dementia, in: M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, S. Duchesne (Eds.), Medical Image Computing and Computer Assisted Intervention - MICCAI 2017, Springer International Publishing, Cham, 2017, pp. 30–37.

[121] M. Bilgel, B. M. Jedynak, Predicting time to dementia using a quantitative template of disease progression, Alzheimer's Dement. Diagnosis, Assess. Dis. Monit. 11 (2019) 205–215.

[122] M. Dawson, H.-G. Müller, Dynamic Modeling of Conditional Quantile Trajectories, with Application to Longitudinal Snippet Data, J. Am. Stat. Assoc. (2017).

[123] D. Goyal, Z. Syed, J. Wiens, Clinically Meaningful Comparisons Over Time: An Approach to Measuring Patient Similarity based on Subsequence Alignment, ArXiv (2018).

[124] S. Liu, S. Liu, S. Pujol, R. Kikinis, D. Feng, M. Fulham, W. Cai, Morphometry-Based Longitudinal Neurodegeneration Simulation with MR Imaging (2015).

[125] H. Chenhui, X. Hua, P. Thompson, G. Fakhri, Q. Li, Inferring sources of dementia progression with network diffusion model, Lect. Notes Comput. Sci. 8679 (2014) 42–49.

[126] K. Franke, G. Ziegler, S. Klöppel, C. Gaser, Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters, Neuroimage 50 (2010) 883–892.

[127] M. Lorenzi, X. Pennec, G. B. Frisoni, N. Ayache, Disentangling normal aging from Alzheimer's disease in structural MR images, Neurobiol Aging 16 (9) (2014) 801.

[128] M. Silver, E. Janousova, X. Hua, P. M. Thompson, G. Montana, Identification of gene pathways implicated in Alzheimer's disease using longitudinal imaging phenotypes with sparse regression., Neuroimage 63 (2012) 1681–1694.

[129] A. Schmidt-Richberg, R. Guerrero, C. Ledig, H. Molina-Abril, A. F. Frangi, D. Rueckert, Multi-stage biomarker models for progression estimation in Alzheimer's disease, in: Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), volume 9123, Springer, Cham, 2015, pp. 387–398. doi:10.1007/978-3-319-19992-4_30.

[130] D. Goyal, D. Tjandra, R. Q. Migrino, B. Giordani, Z. Syed, J. Wiens, Characterizing heterogeneity in the progression of Alzheimer's disease using longitudinal clinical and neuroimaging biomarkers, Alzheimer's Dement. Diagnosis, Assess. Dis. Monit. 10 (2018) 629–637.

[131] R. V. Marinescu, A. Eshaghi, M. Lorenzi, A. L. Young, N. P. Oxtoby, S. Garbarino, S. J. Crutch, D. C. Alexander, DIVE: A spatiotemporal progression model of brain pathology in neurodegenerative disorders, Neuroimage 192 (2019) 166–177.

[132] D. Gamberger, N. Lavrač, S. Srivatsa, R. E. Tanzi, P. M. Doraiswamy, Identification of clusters of rapid and slow decliners among subjects at risk for Alzheimer's disease, Sci. Rep. 7 (2017) 6763.

[133] S. Rathore, M. Habes, M. A. Iftikhar, A. Shacklett, C. Davatzikos, A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages, 2017. doi:10.1016/j.neuroimage.2017.03.057.

[134] M. Huang, W. Yang, Q. Feng, W. Chen, Longitudinal measurement and hierarchical classification framework for the prediction of Alzheimer's disease, Sci. Rep. 7 (2017).

[135] C. R. Jack, D. S. Knopman, W. J. Jagust, R. C. Petersen, M. W. Weiner, P. S. Aisen, L. M. Shaw, P. Vemuri, H. J. Wiste, S. D. Weigand, T. G. Lesnick, V. S. Pankratz, M. C. Donohue, J. Q. Trojanowski, Tracking pathophysiological processes in Alzheimer's disease: An updated hypothetical model of dynamic biomarkers, Lancet Neurol. 12 (2013) 207–216.

[136] M. Gönen, E. Alpaydın, Multiple Kernel Learning Algorithms, J. Mach. Learn. Res. 12 (2011) 2211–2268.

[137] S. Minhas, A. Khanum, F. Riaz, S. A. Khan, A. Alvi, Predicting progression from mild cognitive impairment to Alzheimer's disease using autoregressive modelling of longitudinal and multimodal biomarkers, IEEE J. Biomed. Heal. Informatics 22 (2018) 818–825.

[138] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M. O. Habert, M. Chupin, H. Benali, O. Colliot, Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database, Neuroimage 56 (2011) 766–781.

[139] A. F. Mendelson, M. A. Zuluaga, M. Lorenzi, B. F. Hutton, S. Ourselin, Selection bias in the reported performances of AD classification pipelines, NeuroImage Clin. 14 (2017) 400–416.

[140] J. Samper-González, N. Burgos, S. Bottani, S. Fontanella, P. Lu, A. Marcoux, A. Routier, J. Guillon, M. Bacci, J. Wen, A. Bertrand, H. Bertin, M.-O. Habert, S. Durrleman, T. Evgeniou, O. Colliot, Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data, Neuroimage 183 (2018) 504–521.

[141] B. T. Wyman, D. J. Harvey, K. Crawford, M. A. Bernstein, O. Carmichael, P. E. Cole, P. K. Crane, C. Decarli, N. C. Fox, J. L. Gunter, D. Hill, R. J. Killiany, C. Pachai, A. J. Schwarz, N. Schuff, M. L. Senjem, J. Suhy, P. M. Thompson, M. Weiner, C. R. Jack, Standardization of analysis sets for reporting results from ADNI MRI data, Alzheimer's Dement. 9 (2013) 332–337.

[142] R. J. Bateman, C. Xiong, T. L. Benzinger, A. M. Fagan, A. Goate, N. C. Fox, D. S. Marcus, N. J. Cairns, X. Xie, T. M. Blazey, D. M. Holtzman, A. Santacruz, V. Buckles, A. Oliver, K. Moulder, P. S. Aisen, B. Ghetti, W. E. Klunk, E. McDade, R. N. Martins, C. L. Masters, R. Mayeux, J. M. Ringman, M. N. Rossor, P. R. Schofield, R. A. Sperling, S. Salloway, J. C. Morris, Clinical and Biomarker Changes in Dominantly Inherited Alzheimer's Disease, N. Engl. J. Med. 367 (2012) 795–804.

[143] X. Liu, Methods and Applications of Longitudinal Data Analysis, 2016. doi:10.1016/c2013-0-13082-6.

[144] R. Y. Lo, W. J. Jagust, Predicting missing biomarker data in a longitudinal study of Alzheimer disease., Neurology 78 (2012) 1376–1382.

[145] S. Adhikari, F. Lecci, J. T. Becker, B. W. Junker, L. H. Kuller, O. L. Lopez, R. J. Tibshirani, High-dimensional longitudinal classification with the multinomial fused lasso, Stat. Med. 38 (2019) 2184–2205.

[146] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, J. Tohka, Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects, Neuroimage 104 (2015) 398–412.

[147] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P.-M. Agapow, M. Zietz, M. M. Hoffman, W. Xie, G. L. Rosen, B. J. Lengerich, J. Israeli, J. Lanchantin, S. Woloszynek, A. E. Carpenter, A. Shrikumar, J. Xu, E. M. Cofer, C. A. Lavender, S. C. Turaga, A. M. Alexandari, Z. Lu, D. J. Harris, D. DeCaprio, Y. Qi, A. Kundaje, Y. Peng, L. K. Wiley, M. H. S. Segler, S. M. Boca, S. J. Swamidass, A. Huang, A. Gitter, C. S. Greene, Opportunities and obstacles for deep learning in biology and medicine, J. R. Soc. Interface 15 (2018).

[148] C. R. Jack, D. A. Bennett, K. Blennow, M. C. Carrillo, B. Dunn, S. B. Haeberlein, D. M. Holtzman, W. Jagust, F. Jessen, J. Karlawish, E. Liu, J. L. Molinuevo, T. Montine, C. Phelps, K. P. Rankin, C. C. Rowe, P. Scheltens, E. Siemers, H. M. Snyder, R. Sperling, C. Elliott, E. Masliah, L. Ryan, N. Silverberg, NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease, Alzheimer's Dement. 14 (2018) 535–562.

[149] M. Talo, O. Yildirim, U. B. Baloglu, G. Aydin, U. R. Acharya, Convolutional neural networks for multi-class brain disease detection using MRI images, Comput. Med. Imaging Graph. 78 (2019).

[150] M. Talo, U. B. Baloglu, Ö. Yıldırım, U. Rajendra Acharya, Application of deep transfer learning for automated brain abnormality classification using MR images, Cogn. Syst. Res. 54 (2019) 176–188.

[151] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, D. Feng, Early diagnosis of Alzheimer's disease with deep learning, in: 2014 IEEE 11th Int. Symp. Biomed. Imaging, ISBI 2014, Institute of Electrical and Electronics Engineers Inc., 2014, pp. 1015–1018. doi:10.1109/isbi.2014.6868045.

[152] T. Jo, K. Nho, A. J. Saykin, Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data, Front. Aging Neurosci. 11 (2019).

[153] C. A. Lane, J. Hardy, J. M. Schott, Alzheimer's disease, 2018. doi:10.1111/ene.13439.

[154] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, H. M. Arrighi, Forecasting the global burden of Alzheimer's disease, Alzheimer's Dement. 3 (2007) 186–191.

[155] R. Schmidt, Comparison of Magnetic Resonance Imaging in Alzheimer's Disease, Vascular Dementia and Normal Aging, Eur Neurol 32 (1992) 164–169.

[156] R. A. Rissman, J. Q. Trojanowski, L. M. Shaw, P. S. Aisen, Longitudinal plasma amyloid beta as a biomarker of Alzheimer's disease., J. Neural Transm. 119 (2012) 843–850.

[157] K. Blennow, H. Hampel, M. Weiner, H. Zetterberg, Cerebrospinal fluid and plasma biomarkers in Alzheimer's disease, Nat. Rev. Neurol. 6 (2010) 131–144.

[158] G. Chételat, B. Desgranges, V. De la Sayette, F. Viader, F. Eustache, J. C. Baron, Mild cognitive impairment: Can FDG-PET predict who is to rapidly convert to Alzheimer's disease?, Neurology 60 (2003) 1374–1377.

[159] N. K. Al-Qazzaz, S. H. B. M. Ali, S. A. Ahmad, K. Chellappan, M. S. Islam, J. Escudero, Role of EEG as biomarker in the early detection and classification of dementia, ScientificWorldJournal. 2014 (2014) 906038.

[160] S. Bhat, U. Rajendra Acharya, N. Dadmehr, H. Adeli, Clinical neurophysiological and automated EEG-based diagnosis of the Alzheimer's disease, Eur. Neurol. 74 (2015) 202–210.

[161] N. Houmani, F. Vialatte, E. Gallego-Jutglà, G. Dreyfus, V. H. Nguyen-Michel, J. Mariani, K. Kinugawa, Diagnosis of Alzheimer's disease with electroencephalography in a differential framework, PLoS One 13 (2018).

[162] S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga, L. Beckett, The Alzheimer's disease neuroimaging initiative, Neuroimaging Clin. N. Am. 15 (2005) 869–877.

[163] J. W. Ashford, F. A. Schmitt, Modeling the time-course of Alzheimer dementia., Curr. Psychiatry Rep. 3 (2001) 20–28.

Appendix A. Alzheimer's Disease assessment and markers

Alzheimer's Disease (AD) is characterized by a progressive degeneration of the brain and cognitive functions [153]. In the literature, diagnosis of patients is usually divided in three stages [154], although other classifications have been recently proposed (Section Appendix B):

1. Healthy Control or Cognitively Normal (CN), when the patient shows neither signs of the disease nor cognitive problems.
2. Mild Cognitive Impairment (MCI), when the patient shows signs of cognitive impairment. It can be divided into two substages: early MCI and late MCI, differentiating between patients by their degree of cognitive impairment.
3. AD, when the patient is considered to have completely progressed into full-blown dementia.

Figure A.1 shows an MRI axial view of two different patients: one healthy control and the other with AD. We can appreciate the effects of the disease directly on the reduction of cortical thickness, among other visual and physical cues [155].

AD markers

To determine the stage of the disease, various markers describing key pathophysiological processess of AD have been proposed over the years. Markers of the brain provide information for the study of the disease and its screening. AD is characterized by protein amyloid-beta (Aβ) deposition in the brain [156], tau injury, and structural neurodegeneration [135]. Those three indicators precede cognitive impairment, leading to death. For the measurement of these indicators, different markers have been proposed:

1. Brain Aβ deposition in the brain can be detected both in positron emission tomography (PET) imaging [50], and in cerebrospinal fluid (CSF) [49].

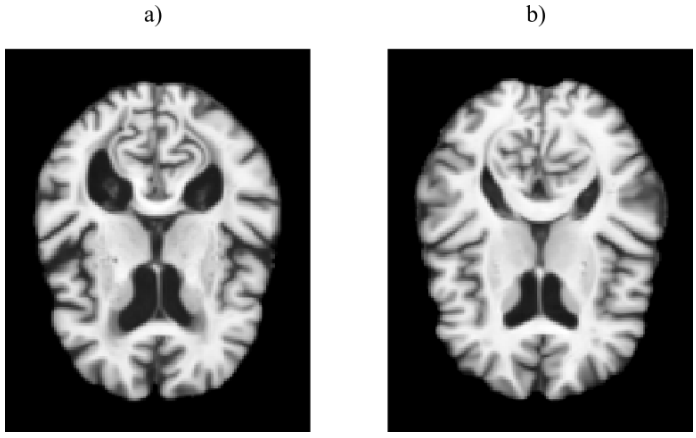

Figure A.1: Axial view of MRI scan for AD (left) and CN (right) patients. Images from ADNI dataset, registered to a common template.

2. Tau injury and dysfunction caused by tau and p-tau plaques, found in tau-PET imaging and CSF [49, 157].

3. Neurodegeneration provoked by tau injury. It can be observed in structural magnetic resonance imaging (MRI) [51] and in fludeoxyglucose (FDG)-PET imaging [158].

4. Memory and cognition, measured by cognitive tests.

The main screening tool for clinical assessment of AD is the clinical interview between the patient and the doctor, where the severity of the cognitive problems of the patient can be assessed, followed by a cognitive physical examination to capture the aforementioned markers and assess the presence of the disease [153].

Apart from the aforementioned markers and imaging techniques, resting-state electroencephalography (EEG) signals have also been proposed for AD assessment [159, 160]. However, they are not as widely used as image-based examination, as EEG cannot be used to observe specific processes in the brain and they only show changes in brain activity, which could be caused by other pathologies. A review on EEG methods for AD can be found in [161].

Longitudinal marker dynamics and disease model

The previous markers can be studied and modelled longitudinally. Modelling their trajectories and progression can give us more insight on how they change and interact. For example, longitudinal data analysis on MRI allows us to calculate the rate of change of specific brain structures, such as the dynamics of cortical and hippocampal atrophy.

A widely accepted progression model of AD was proposed by [5]. Their model is based on marker evolution, where each marker progresses from normal values to abnormal values differently. The order of the markers is the presented above: Aβ deposition, followed by tau injury, neurodegeneration and cognition. Empirical data and experiments reviewed in [135] confirm the validity of the model, although other data-driven works do not fully agree with it [39]. Analyzing those markers longitudinally allows us to study both the individual and whole population rate of change, and improve AD progression modelling.

Studies and initiatives

There has been a remarkable number of initiatives to promote using longitudinal data on AD modelling. Availability of patients' longitudinal data is key to study the progression of the disease. [6] presented a review of available longitudinal AD biomarker datasets, finding that more efforts are needed to increase the follow-up duration, increase the population sizes and standardize the acquisition methods. One of the largest studies is the Alzheimer's Disease Neuroimaging Initiative (ADNI) [162], a multimodal, ongoing longitudinal study with hundreds of enrolled subjects, gathering imaging data, cognitive scores, blood and CSF markers. To unify and share the available data, the Alzheimer's Association has created The Global Alzheimer's Association Interactive Network (GAAIN) [2] to share data between independent studies and build collaborations

[2] http://www.gaain.org

to create and explore large, heterogeneous cohorts.

Initiatives to stimulate research on the field have also been proposed, such as the MIRIAD challenge [70], The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) Challenge[3] or Quantitative Templates for the Progression of Alzheimer's disease (QT-PAD)[4]. These challenges define a fixed subset of available data, making it easier to compare results, share methods and ensure reproducibility.

Appendix B. NIA-AA research framework: new biological definition on AD

A new unified research framework for a biological definition of the disease was recently published by the National Institute on Aging and the Alzheimer's Association [148]. This approach defines AD as a biological construct based on markers, rather than clinical symptoms of the disease such as cognitive impairments. The framework is flexible enough for the introduction of additional markers, if needed.

The framework groups markers in three categories: Aβ deposition, pathologic tau, and neurodegeneration. This is represented as the AT(N) system, where each category can be binarized using a cut point into normal/abnormal (-/+). For each category:

- **A:** Aβ markers determine if a patient is in the Alzheimer's continuum, showing pathological changes but still not presenting the disease.
- **T:** tau deposition markers indicate whether a patient who is in the Alzheimer's continuum has the disease.

[3] `https://tadpole.grand-challenge.org`
[4] `http://www.pi4cs.org/qt-pad-challenge`

- **(N):** Neurodegeneration markers show structural changes in the brain that can be product of AD, but are not specific to the disease (and thus is placed in parenthese).

The flexibility of the framework allows working with missing biomarker values, which is a valuable trait for a longitudinal study. We found no work (within the scope of this review) using this new biological definition on AD. The reasons could be the recentness of the framework's publication, and the need for multimodal data in a longitudinal setting, which is not as available as single modality MRI. However, we expect future studies to use this framework, as it offers clear advantages for longitudinal analysis: for example, being able to directly compare different stages of progression between patients, or extend the framework with markers that capture longitudinal progression.

Appendix C. Challenges in longitudinal data

Longitudinal data are composed of sequential data acquisitions for subjects over a period of time. This contrasts with cross-sectional studies, which focus on single acquisitions per subject. Here, we describe the main characteristics and analysis challenges that arise while dealing with longitudinal data. In Section 6 we outline strategies to overcome some of them.

Two sources of variability can be defined for a longitudinal study in a cohort of subjects: the inter-subject variability, i.e., the differences between observations of different subjects, and the intra-subject variability, i.e., the differences between observations of a same subject, which tend to be highly correlated compared to the former. Those two sources of variation give valuable information about the progression of the disease between- and within- subjects. In cross-sectional studies, those two variabilities are non-separable: given two samples of different subjects, it is not possible to know to what extent their variation is due to inter-subject variability or to the different stages of the disease. Adding longitudinal samples for each subject allows us to distinguish between those two

variabilities, improving our understanding of the disease [13].

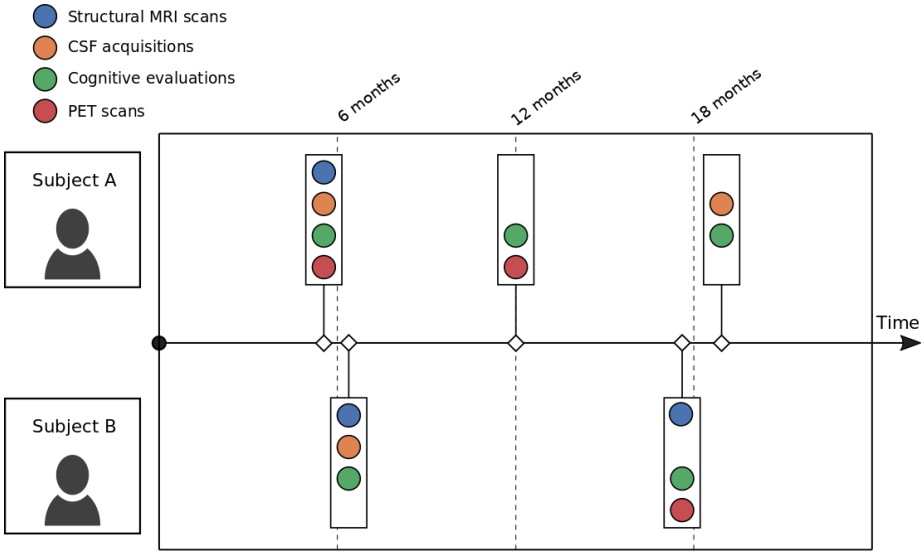

Figure C.2: Longitudinal representation of data acquisitions for two patients.

Figure C.2 shows an example of a longitudinal study for two subjects, with multiple data modalities, over a fixed span of time. It illustrates some of the challenges that can appear in a longitudinal, multimodal data study:

1. Each subject can have a different number of acquisitions, leading to an unbalanced data problem. In the figure, Patient B missed the 12th month acquisition for some reasons.
2. There can be missing data due to missing acquisitions from some modalities. In the figure, only patient A at the 6-months follow-up has all the acquisitions.
3. Data are not necessarily acquired at the same time point for the different subjects.
4. Time spacing between follow-ups can be variable, even within a single subject.

Another problem, not shown in the figure, is that different patients can be at different stages of the disease at a given time point. Reference time to measure progression remains an open issue in the field [163].

Protocols of data acquisition try to palliate these problems, but in a clinical setting, this is very difficult to achieve: sometimes patients miss their scheduled screening session and data cannot be gathered. Other patients might drop out from the study for a variety of reasons, such as disease severity or moving out of the city/country, and in other cases, data of a given time point could need to be discarded because of quality problems. For these reasons, most of the available longitudinal data is unbalanced.

All studies should define their policy on this issue, either by selecting only subjects with no missing data in their studies, or by defining a method to handle the problem. Popular methods for missing data in longitudinal studies are detailed in [14].