



HAL
open science

Multilevel Survival Modeling with Structured Penalties for Disease Prediction from Imaging Genetics data

Pascal Lu, Olivier Colliot

► **To cite this version:**

Pascal Lu, Olivier Colliot. Multilevel Survival Modeling with Structured Penalties for Disease Prediction from Imaging Genetics data. *IEEE Journal of Biomedical and Health Informatics*, 2022, 26 (2), pp.798-808. 10.1109/JBHI.2021.3100918 . hal-03311509v2

HAL Id: hal-03311509

<https://inria.hal.science/hal-03311509v2>

Submitted on 14 Aug 2021

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

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

Multilevel Survival Modeling with Structured Penalties for Disease Prediction from Imaging Genetics data

Pascal Lu and Olivier Colliot, *Member, IEEE*, for the Alzheimer's Disease Neuroimaging Initiative

Abstract—This paper introduces a framework for disease prediction from multimodal genetic and imaging data. We propose a multilevel survival model which allows predicting the time of occurrence of a future disease state in patients initially exhibiting mild symptoms. This new multilevel setting allows modeling the interactions between genetic and imaging variables. This is in contrast with classical additive models which treat all modalities in the same manner and can result in undesirable elimination of specific modalities when their contributions are unbalanced. Moreover, the use of a survival model allows overcoming the limitations of previous approaches based on classification which consider a fixed time frame. Furthermore, we introduce specific penalties taking into account the structure of the different types of data, such as a group lasso penalty over the genetic modality and a ℓ_2 -penalty over the imaging modality. Finally, we propose a fast optimization algorithm, based on a proximal gradient method.

The approach was applied to the prediction of Alzheimer's disease (AD) among patients with mild cognitive impairment (MCI) based on genetic (single nucleotide polymorphisms - SNP) and imaging (anatomical MRI measures) data from the ADNI database. The experiments demonstrate the effectiveness of the method for predicting the time of conversion to AD. It revealed how genetic variants and brain imaging alterations interact in the prediction of future disease status. The approach is generic and could potentially be useful for the prediction of other diseases.

Index Terms—Multilevel model, survival model, Cox Proportional Hazards, Group Lasso penalty, imaging genetics, Alzheimer's Disease

I. INTRODUCTION

BETTER understanding, diagnosing and predicting diseases requires integration of multimodal data. In particular, the joint use of medical imaging and genetic data, a field often called imaging genetics or radiogenomics, holds major

The research leading to these results has received funding from the French government under management of *Agence Nationale de la Recherche* as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute) and reference ANR-10-IAIHU-06 (Agence Nationale de la Recherche-10-IA Institut Hospitalo-Universitaire-6) and from the INRIA Project Lab Program (project Neuromarkers).

P. Lu is with Sorbonne Université, Inserm, CNRS, Paris Brain Institute (ICM), AP-HP - Hôpital Pitié-Salpêtrière, INRIA Paris, ARAMIS project-team, Boulevard de l'hôpital, F-75013, Paris, France (e-mail: pascal.lu@outlook.com).

O. Colliot is with Sorbonne Université, Inserm, CNRS, Paris Brain Institute (ICM), AP-HP - Hôpital Pitié-Salpêtrière, INRIA Paris, ARAMIS project-team, Boulevard de l'hôpital, F-75013, Paris, France (e-mail: olivier.colliot@sorbonne-universite.fr).

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.ucla.edu>). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and provided data but did not participate in analysis or writing of this report.

promises for precision medicine [37]. This is for instance the case in cancer and Alzheimer's disease, as highlighted by a recent review [37]. Many works in this area have studied the statistical association between genetic variants and imaging characteristics [22]. In particular, a large number of papers studied the relationship between genetic and imaging data by considering that an imaging phenotype can be explained by a sum of effects from genetic variants. These multivariate approaches use partial least squares [31], sparse canonical correlation analysis [32], sparse regularized linear regression with a ℓ_1 -penalty [24], group lasso penalty [26], [25], or Bayesian modelling that links genetic variants to imaging variables and imaging variables to the disease status [23].

Another interesting problem deals with the combination of genetic and medical imaging data for disease prediction. For instance, there is a major interest in predicting the future occurrence of Alzheimer's disease (AD) in patients with mild cognitive impairment (MCI) (see [2] for a recent review on that topic). The group of patients with MCI is heterogeneous, some being more likely than others to convert to AD [36]. Beyond this paper, other interesting tasks include prediction of dementia in patients with Parkinson's disease [65] or predicting drug response in different diseases such as mental disorders [66].

Machine learning methods have been proposed to build predictors from multimodal data. However, challenging issues are high-dimensionality data, small number of observations, the heterogeneous nature of data, and the different weights that each modality may have. A framework that is commonly used to combine heterogeneous data is multiple kernel learning (MKL) [19], [30], where each modality is represented by a kernel. The decision function and weights for the kernels are simultaneously learnt. Moreover, the group lasso [15], [16] is a way to integrate structure inside data. However, the standard ℓ_1 -MKL and group lasso may eliminate modalities that have a weaker contribution. To overcome this problem, different papers have proposed to use a $\ell_{1,p}$ -penalty [20] to combine optimally different modalities [27], [28], [29]. However, these approaches do not consider potential interactions between genetic and imaging data. Moreover, they put genetic and imaging data on the same level, although these data do not provide the same type of information: genetic data modulates disease risk while imaging data characterizes abnormalities at a specific time.

Most existing approaches to predict the future occurrence of the disease define the problem as a classification task at fixed

time [38], [49], [47]: one chooses a date (e.g. 36 months) and aims to discriminate between patients who will convert before that date and those who will not. However, choosing a date is arbitrary and suboptimal as non-converters may in fact be patients who will convert shortly after this date. It would be more interesting to be able to estimate the date of conversion. One can aim to solve multiple classification problems, each for a different data, but then there is no guarantee that the different predictions will be consistent. Survival models provide a more principled way to estimate the conversion date. Unlike multiple fixed-time classifications, survival models ensure that the conversion date is monotonous. Survival models (in particular, the Cox Proportional Hazards model) have been applied for combining multimodal data [39] and for predicting the conversion to AD [41], [40], [43], [67], [3]. In [43], the authors show that the combination of both genetic and neuroimaging modalities enhances the prediction. However, as the classification approaches mentioned above, they use an additive framework in which each modality is put at the same level. Beyond the standard Cox model, more sophisticated survival approaches have been proposed for predicting progression to AD among MCI patients, such as survival forests [4], combining Cox and linear mixed-effects models [5], combining survival forests and the Cox model [1] or including radiomics-based biomarkers into a Cox model [7]. Deep learning approaches have also been proposed for survival analysis of progression to AD. This includes deep survival analysis which extends a regularized multi-layer perceptron to account for censored data [8] and temporal convolutional networks [9]. However, these different techniques do not combine imaging and genetic data.

In this paper, we propose a multilevel survival model for disease prediction from imaging and genetic data. The original multilevel setting allows modeling interactions between genes and imaging variables in disease prediction. Specifically, we considered that the decision function parameters learnt from imaging data could be modulated by the genetic data of each subject. We integrated specific penalties which are adapted to each modality: a ℓ_2 -penalty on the imaging modality, and a group lasso penalty over the genetic modality which models the mapping of variants into genes. The approach was applied to prediction of AD among patients with MCI.

Preliminary versions of this work were presented at a workshop [11] and a conference [12]. The present paper provides a more thorough mathematical description of the model and more extensive experiments.

II. MODEL SET-UP

We present the multilevel survival model for disease prediction from imaging and genetic data. We first introduce the multilevel survival model (section II-A), then the penalties (section II-B) and finally a fast optimization procedure based on a proximal gradient method (section II-C).

A. Multilevel Survival model

Let $\mathbf{x}_G \in \mathbb{R}^{|\mathcal{G}|}$ represent the genetic data, $\mathbf{x}_I \in \mathbb{R}^{|\mathcal{I}|}$ be the imaging data.

For all patients, the entry in the study is the starting point of the survival model, and the time at which they progressed to a diseased state is the conversion date. In our application to Alzheimer's Disease, the patients are MCI at the entry in the study and the conversion date is the time at which they progressed to AD.

For each patient i , T_i^* is his real conversion date (which may not be observed), C_i is the date of his final visit and $T_i = \min(T_i^*, C_i)$ is the observed duration. We define $\delta_i = \mathbf{1}_{\{T_i^* \leq C_i\}}$, which indicates if the conversion has occurred.

The conversion date T can be seen as a random variable with cumulative distribution function

$$F : t \mapsto \mathbb{P}\{T < t | \mathbf{x}_G, \mathbf{x}_I\} = 1 - S(t) \quad (1)$$

where the survival function S is defined as:

$$\begin{aligned} S(t) &= \mathbb{P}\{T \geq t | \mathbf{x}_G, \mathbf{x}_I\} \\ &= \exp\left(-\int_0^t h(u | \mathbf{x}_G, \mathbf{x}_I) du\right) \end{aligned} \quad (2)$$

and h is the hazard function. It represents the instantaneous rate of occurrence of the conversion.

Based on the Cox proportional hazards assumption, we propose the multilevel framework:

$$h(t | \mathbf{x}_G, \mathbf{x}_I) = h_0(t) e^{\beta(\mathbf{x}_G)^\top \mathbf{x}_I} \quad (3)$$

where h_0 is the baseline hazard function describing the risks for individuals whose covariates are null and $\beta(\mathbf{x}_G)$ is the parameter vector depending on genetic data \mathbf{x}_G .

Unlike in the classical Cox model, where imaging and genetic data are put at the same level, this is a multilevel model for which the parameter vector $\beta(\mathbf{x}_G)$ depends on genetic data \mathbf{x}_G .

β is assumed to be an affine function of genetic data \mathbf{x}_G :

$$\beta(\mathbf{x}_G) = \mathbf{W}\mathbf{x}_G + \beta_I \quad (4)$$

where $\mathbf{W} \in \mathbb{R}^{|\mathcal{I}| \times |\mathcal{G}|}$ and $\beta_I \in \mathbb{R}^{|\mathcal{I}|}$.

Then,

$$h(t | \mathbf{x}_G, \mathbf{x}_I) = h_0(t) e^{\mathbf{x}_G^\top \mathbf{W}^\top \mathbf{x}_I + \beta_I^\top \mathbf{x}_I} \quad (5)$$

Finally, the survival function is:

$$S(t | \mathbf{x}_G, \mathbf{x}_I) = (S_0(t))^{e^{\mathbf{x}_G^\top \mathbf{W}^\top \mathbf{x}_I + \beta_I^\top \mathbf{x}_I}} \quad (6)$$

where $S_0(t) = \exp\left(-\int_0^t h_0(u) du\right)$.

Relationships between parameters are summarized in Figure 1.

From the dataset $\mathcal{D} = \{(\mathbf{x}_G^i, \mathbf{x}_I^i, T_i, \delta_i), i = 1, \dots, N\}$ where the covariates $\mathbf{x}_G, \mathbf{x}_I$ and the cross-product covariates $\mathbf{x}_G \mathbf{x}_I^\top$ are centered and normalized, the negative partial log-likelihood is:

$$\begin{aligned} \mathcal{L}(\mathbf{W}, \beta_I) &= -\frac{1}{N} \sum_{i=1}^N \delta_i (\mathbf{x}_G^i)^\top \mathbf{W}^\top \mathbf{x}_I^i + \beta_I^\top \mathbf{x}_I^i \\ &\quad - \log \sum_{j \in \mathcal{R}(T_i)} e^{\mathbf{x}_G^j \mathbf{W}^\top \mathbf{x}_I^j + \beta_I^\top \mathbf{x}_I^j} \end{aligned} \quad (7)$$

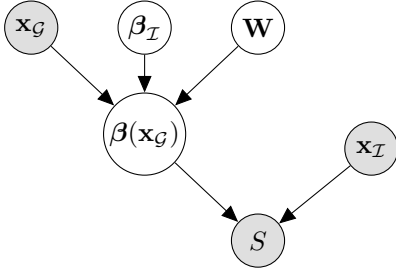


Fig. 1: The survival is a function of imaging data \mathbf{x}_I and the parameter $\beta(\mathbf{x}_G)$ (which is computed from genetic data \mathbf{x}_G)

where $\mathcal{R}(T_i)$ represents the set of patients j such that $T_j \geq T_i$.

The baseline survival function S_0 is usually estimated using a non-parametric estimator, such as the Kaplan-Meier estimator [21], on the dataset $\{(T_i, \delta_i), i = 1, \dots, N\}$.

B. Penalties

Genetic data can be represented as a sequence of single-polymorphism nucleotides (SNP) counted by minor allele. Each SNP can belong to zero, one or several genes. Genes participate in the production of proteins that interact inside pathways. SNPs are grouped by genes, and we designed a penalty to enforce regularity inside genes and sparsity between genes. Similarly, it is possible to group SNPs by pathways instead of genes. Because some SNPs may belong to several genes, the group lasso with overlap penalty [17] is more suitable. An overlap expansion is performed to deal with this penalty. From the subject's feature vector $\mathbf{x} \in \mathbb{R}^{|\mathcal{G}|}$, a new feature vector is created $\tilde{\mathbf{x}} = (\mathbf{x}_{\mathcal{G}_1}^\top, \dots, \mathbf{x}_{\mathcal{G}_L}^\top)^\top \in \mathbb{R}^{\sum_{\ell=1}^L |\mathcal{G}_\ell|}$, defined by the concatenation of copies of the genetic data restricted by group \mathcal{G}_ℓ . We also performed the same expansion on \mathbf{W} to obtain $\tilde{\mathbf{W}} \in \mathcal{M}_{|\mathcal{I}|, (\sum_{\ell=1}^L |\mathcal{G}_\ell|)}(\mathbb{R})$. For the matrix \mathbf{W} , we also used the group lasso with overlap penalty where groups are (brain regions, genes).

For imaging variables, we used the ridge penalty:

$$\Omega_I(\beta_I) = \|\beta_I\|_2^2 \quad (8)$$

Finally, the overall penalty Ω is defined by:

$$\Omega(\tilde{\mathbf{W}}, \beta_I) = \lambda \sum_{i=1}^{|\mathcal{I}|} \sum_{\ell=1}^L \theta_{\mathcal{G}_\ell} \left\| \tilde{\mathbf{W}}_{i, \mathcal{G}_\ell} \right\|_2 + \lambda_I \|\beta_I\|_2 \quad (9)$$

where $\lambda > 0, \lambda_I > 0$ are the hyperparameters.

Group weighting for groups (brain regions, genes) is based on the number of SNPs in the gene: for the gene \mathcal{G}_ℓ , the weight $\theta_{\mathcal{G}_\ell} = \sqrt{|\mathcal{G}_\ell|}$ ensures that the penalty term is of the order of the number of parameters of the group. The hyperparameter λ influences the number of groups that are chosen by the model. In particular, a group \mathcal{G}_ℓ is considered in the model during the first iteration if $\left\| \nabla_{\mathbf{W}_{i, \mathcal{G}_\ell}} R_N(\mathbf{0}) \right\|_2 > \lambda \theta_{\mathcal{G}_\ell}$. An upper bound for λ is given by this inequality.

C. Optimization

The parameters \mathbf{W}, β_I are obtained by finding a minimum of $\mathcal{L}(\mathbf{W}, \beta_I) + \Omega(\mathbf{W}, \beta_I)$. The classical approach for dealing

with this problem is to use a proximal gradient descent on the convex set defined by Ω [14], [18].

The cross-product covariates $\mathbf{x}_G \mathbf{x}_I^\top$ and the matrix \mathbf{W} are flattened and transformed into a vector. A vector $\gamma = (\text{flatten}(\mathbf{W}), \beta_I)$ containing the coefficients of \mathbf{W}, β_I is created. The matrix \mathbf{W} can be recreated by unflattening γ . A proximal gradient descent updates the vector γ at each step. Algorithm 1 describes the whole procedure. This algorithm stops when the stopping criterion is met

$$\left| (\mathcal{L} + \Omega)(\mathbf{W}, \beta_I) - (\mathcal{L} + \Omega)(\widehat{\mathbf{W}}, \widehat{\beta}_I) \right| < \eta |(\mathcal{L} + \Omega)(\mathbf{W}, \beta_I)| \quad (10)$$

where η is a tolerance threshold. The code is made available online ¹.

Algorithm 1: Optimization procedure

```

1 Input: dataset  $\mathcal{D}$ ,  $\delta = 0.5$ ,  $\varepsilon_0 = 0.01$ ,  $\eta = 10^{-5}$ ;
2 Initialization:  $\mathbf{W} = \mathbf{0}$ ,  $\beta_I = \mathbf{0}$ , converged = False
3 while not(converged) do
4    $\gamma = (\text{flatten}(\mathbf{W}), \beta_I)$ ;
5    $\omega = \gamma - \varepsilon \nabla \mathcal{L}$ ;
6   for  $(i, \ell) \in \llbracket 1, |\mathcal{I}| \rrbracket \times \llbracket 1, L \rrbracket$  do
7      $\widehat{\mathbf{W}}_{\mathcal{G}_\ell, i} = \max \left( 0, 1 - \frac{\varepsilon \lambda_{\mathcal{G}} \theta_{\mathcal{G}_\ell}}{\|\omega_{\mathcal{G}_\ell + i|\mathcal{G}_\ell}\|_2} \right) \omega_{\mathcal{G}_\ell + i|\mathcal{G}_\ell}$ ;
8   end
9    $\widehat{\beta}_I = \frac{\omega_{\mathcal{I} + |\mathcal{G}| |\mathcal{I}|}}{1 + 2\varepsilon \lambda_I}$ ;
10  if  $(\mathcal{L} + \Omega)(\widehat{\mathbf{W}}, \widehat{\beta}_I) > (\mathcal{L} + \Omega)(\mathbf{W}, \beta_I)$  then
11     $\varepsilon = \delta \varepsilon$ 
12  else
13     $\mathbf{W} = \widehat{\mathbf{W}}$ ,  $\beta_I = \widehat{\beta}_I$ ,  $\varepsilon = \varepsilon_0$ ;
14  end
15  converged =  $\left| 1 - \frac{(\mathcal{L} + \Omega)(\widehat{\mathbf{W}}, \widehat{\beta}_I)}{(\mathcal{L} + \Omega)(\mathbf{W}, \beta_I)} \right| < \eta$ 
16 end

```

III. EXPERIMENTS AND RESULTS

We applied our approach to the prediction of the date at which patients with mild cognitive impairment (MCI) progressed to Alzheimer's disease (AD).

A. Simulated data

We draw $N = 1000$ subjects. For subject i , we draw

- 10 SNPs grouped in 5 genes $(x_1, x_2), (x_3, x_4), (x_5, x_6), (x_7, x_8), (x_9, x_{10})$ with $x_j \sim \text{Rand}(\{0, 1, 2\})$
- 10 brain regional measures y_1, \dots, y_{10} with $y_j \sim \text{Rand}([0, 20])$
- Compute $Z = \sum_{i,j} M_{i,j} x_i y_j$ where M is a weight matrix.

We chose a matrix of rank 1 in our experiment shown in Figure 2a.

¹<https://github.com/lupascal/MultilevelSurvival>

- Compute the median time T^* using the survival function

$$S(t) = (f(t))^{e^Z}$$

with $f : t \mapsto \exp(-t)$.

- Draw $\delta \sim \text{Rand}(\{0, 1\})$.
If $\delta = 0$, then draw $T \sim \text{Rand}([0, T^*])$, else $T = T^*$.

It can be noted that the baseline survival function S_0 is not f . We performed a train-test split in order to find the best hyperparameter λ_W maximising the C-index over the test set. The input and simulated matrices are shown on Figure 2. When comparing the input and the estimated matrices W , we notice that the strongest interactions coefficients are well recovered by the model. Moreover, all the simulated zeros of the W matrix were correctly estimated, suggesting that the model did not estimate spurious interactions. However, for the non-zero coefficients, the simulated and estimated values could differ, this being possibly due to the regularizations introduced by the model. Overall, we believe that the interactions that are estimated by the model are reasonable.

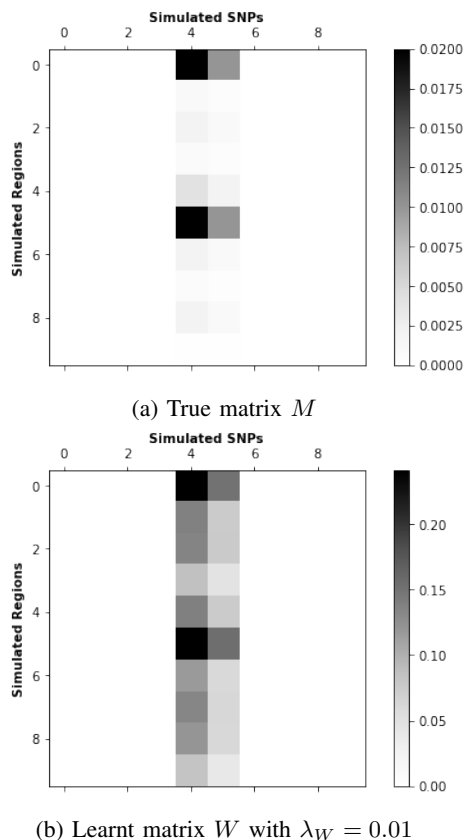


Fig. 2: True matrix M vs learnt matrix W

B. Dataset

The data used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative database (<http://adni.loni.usc.edu>). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial MRI, PET, other biological

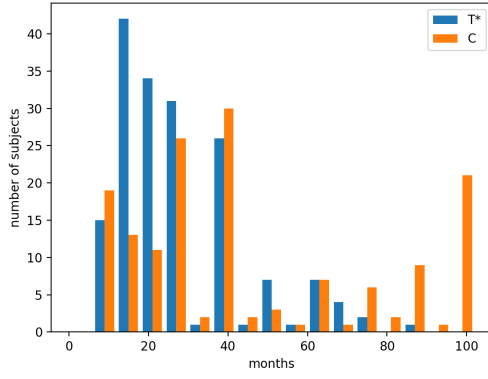
markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. Over 1,650 participants were recruited across North America during the three phases of the study (ADNI1, ADNI GO and ADNI2). Around 400 participants were diagnosed with AD, 900 with MCI and 350 were control subjects. Three main criteria were used to classify the subjects [13]. We considered only the patients who had MCI at baseline. MCI patients had a mini-mental state examination (MMSE) score between 24 and 30 (inclusive) and a clinical dementia rating (CDR) score of 0.5 with a mandatory requirement of the memory box score being 0.5 or greater. The other criteria can be found in [13].

For our experiments, we used the ADNI1 GWAS dataset. It comprises 199 MCI patients at baseline who progressed to AD during the entire study and 147 MCI patients who remained stable throughout the entire study. Genetic data consist of 620,901 genotyped SNPs. To reduce the dimensionality, we selected 1,107 SNPs which correspond to the 44 first top genes related to AD as identified in the AlzGene database, (<http://www.alzgene.org>). As for neuroimaging, we used anatomical MRI data with the following extracted features. Specifically, the segmentation of FreeSurfer gives the average cortical thickness in different regions (70 features) and the volume of subcortical regions (42 features). Therefore, $1,107 \times 112 = 123,984$ parameters must be inferred for W , 112 parameters for β_T .

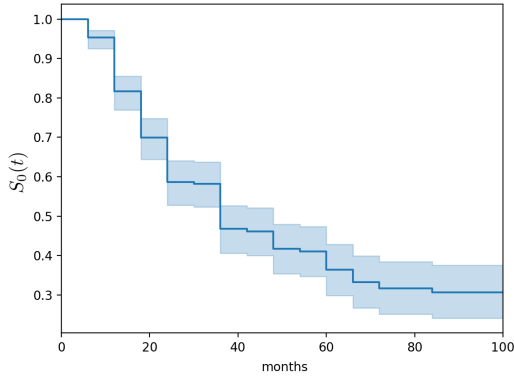
a) *Baseline survival function S_0* : Figure 3(a) displays the distribution of the conversion dates T^* and of the dates of final visits C . Figure 3(b) shows the baseline survival function S_0 computed using the Kaplan-Meier estimator. This represents the probability of staying stable as a function of time and without taking into account the covariates. The variance of the baseline survival function is based on the Greenwood estimator. Figure 3(c) presents the baseline hazard function h_0 computed using the Nelson-Aalen estimator. This represents the instantaneous rate of conversion, not taking into account the covariates. The maximum follow-up date is $\tau_{\text{hor}} = 100$ months; truncation occurs for patients who convert after this date. The median survival time, defined as the smallest survival time for which the survival function is less than or equal to 0.5, is 36 months. One can observe that the baseline hazard function first increases and then decreases, with the exception of months 30, 42 and 48 at which much less patients had a follow-up.

C. Cross-validation

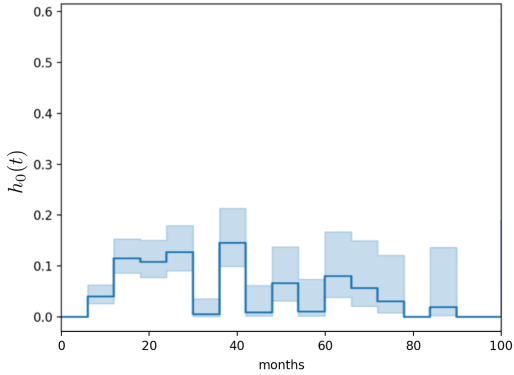
The different models were validated using a 5-fold cross-validation. To determine the optimal hyperparameters in all models without biasing the performance estimates, we use a nested cross-validation. We perform a 5-fold cross-validation, and within each fold, we find the optimal hyperparameters using a 5-fold cross-validation on the training set and taking the hyperparameters that maximize the C-index over the inner test set. For our multilevel model, we maximized the C-index



(a) Histogram of conversion dates T^* and of the dates of final visits C



(b) Kaplan-Meier estimate of the baseline survival function S_0



(c) Nelson-Aalen estimate of the hazard function h_0 (bandwidth = 6 months)

Fig. 3: ADNI1 Dataset: baseline survival function S_0 and hazard function h_0 .

while avoiding solutions such that $\mathbf{W} = 0$. The hyperparameters are optimized in the range

$$\{10^{-4}, 10^{-3}, \dots, 10^{-1}, 1, 10, \dots, 10^4\}$$

It takes 8 minutes to train the model for fixed hyperparameters on a standard laptop (MacBook Pro, Retina, 2014).

D. Comparison to other approaches

We compared the multilevel survival framework to:

- the Cox Proportional Hazards model using only gender and age as input variables
- the Cox Proportional Hazards model with a lasso penalty using only genetic data as input variables
- the Cox Proportional Hazards model with a lasso penalty using genetic data, gender and age as input variables
- the Cox Proportional Hazards model (either without penalty or with the ridge penalty) using only imaging data
- a deep learning approach (deep survival analysis) recently proposed by Nakagawa et al [8] using only imaging data
- the Cox Proportional Hazards using both genetic and imaging data which are combined in an additive framework. In this case, the hazard function is given by:

$$h(t|\mathbf{x}_G, \mathbf{x}_I) = h_0(t)e^{\beta_G^\top \mathbf{x}_G + \beta_I^\top \mathbf{x}_I}$$

- our multilevel model using imaging data and APOE, without the other SNPs

E. Metrics

To assess the quality of the prediction, we use the three following measures:

- the concordance index (or C-index) [35]. It verifies if the model orders the conversion dates in the same order as the ground truth. The C-index can be seen as a generalization of AUC and its range is $[0, 1]$. Typical values are in $[0.55, 0.7]$.
- the integrated Brier score [34], for uncensored data:

$$\text{Brier} = \frac{1}{\tau_{\text{hor}}} \int_0^{\tau_{\text{hor}}} \left[\frac{1}{N} \sum_{i=1}^N \delta_i (\mathbf{1}_{\{T_i^* > t\}} - \hat{S}(t|\mathbf{x}))^2 \right] dt \quad (11)$$

It measures the accuracy of probabilistic predictions. When the Brier score is lower, the model performs better.

- the integrated Area Under Curve

$$\text{iAUC} = \int_0^{\tau_{\text{hor}}} \text{AUC}(t) f(t) dt \quad (12)$$

where τ_{hor} is the maximum follow-up date, f is the probability density function of T and $\text{AUC}(t)$ is the cumulative/dynamic AUC [33].

This AUC has the same role as the classical AUC in classification. The range of the iAUC is $[0, 1]$. When the iAUC is higher, the model performs better.

We also displayed $\text{ROC}(t)$ curves for different time-points. These were computed using using the R-package `survivalROC` with NNE as estimator [68]. They were averaged across test folds².

F. Prediction results

Results are presented on table I. One can observe that models using MRI (with or without genetic) achieved much better performance than those using only genetic, age, gender

²https://scikit-learn.org/stable/auto_examples/model_selection/plot_roc_crossval.html

MODALITY	METHOD & PENALTY	SURVIVAL MODELS		
		C-INDEX	BRIER SCORE	IAUC
Gender + Age	Cox PH model	0.476 ± 0.024	0.237 ± 0.009	0.474 ± 0.032
SNPs only	Cox PH model (lasso penalty)	0.521 ± 0.040	0.166 ± 0.009	0.515 ± 0.031
SNPs + Gender + Age	Cox PH model (lasso penalty)	0.522 ± 0.035	0.166 ± 0.016	0.518 ± 0.031
MRI only	Cox PH model (no penalty)	0.636 ± 0.034	0.190 ± 0.017	0.636 ± 0.050
MRI only	Cox PH model (ridge penalty)	0.671 ± 0.022	0.149 ± 0.008	0.663 ± 0.044
MRI only	Deep Survival Analysis (Nakagawa et al.[8])	0.651 ± 0.034	0.191 ± 0.033	0.658 ± 0.064
MRI + SNPs	Additive Cox PH model (lasso penalty)	0.677 ± 0.020	0.148 ± 0.006	0.680 ± 0.030
MRI + APOE	Multilevel model (ours)	0.653 ± 0.030	0.149 ± 0.015	0.686 ± 0.033
MRI + SNPs	Multilevel model (ours)	0.681 ± 0.018	0.147 ± 0.006	0.686 ± 0.031

TABLE I: Survival results for different modalities and methods (mean value across the test folds \pm standard deviation). Specifically, we compared our multilevel model to a standard (additive) Cox proportional hazards (PH) model, with or without penalty and to a deep survival analysis method. We compared the performances obtained when using MRI and SNPs, MRI and APOE only, MRI only, SNPs only as well as age and gender.

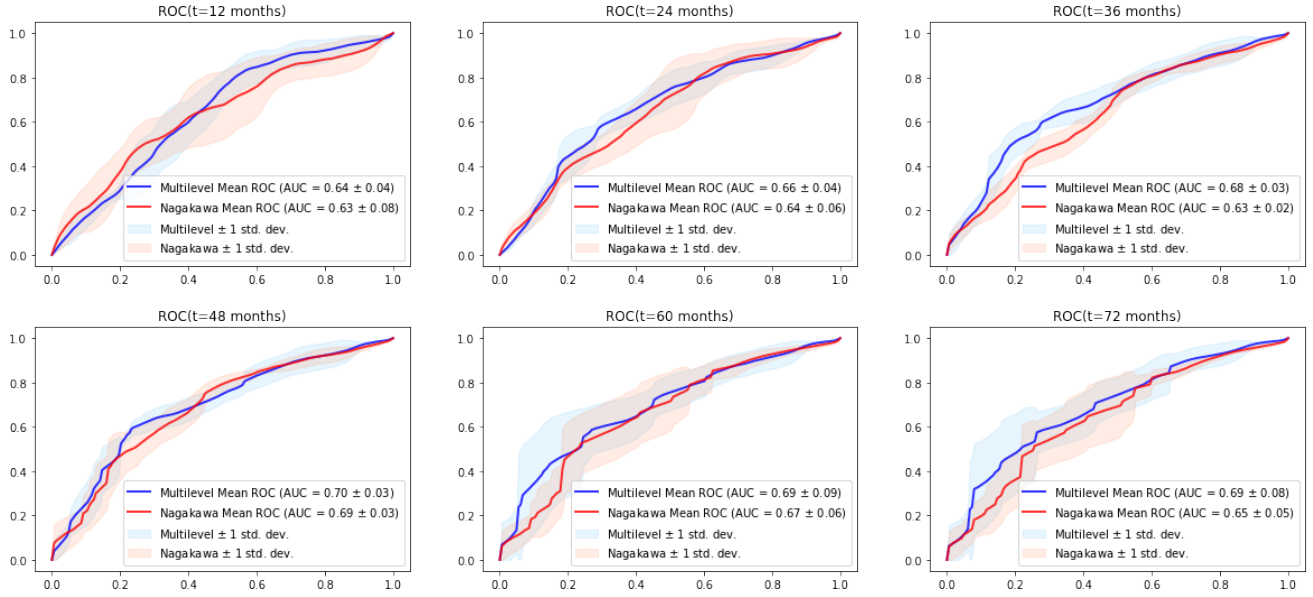


Fig. 4: Mean ROC(t) curves across outer test folds at different time points (12, 24, 36, 48, 60, 72 months) for our model (multilevel) and the deep survival analysis proposed by Nakagawa et al. [8]

or a combination of these three variables. This indicates that MRI is more powerful than genetics, age or gender for predicting progression to AD in MCI patients. One can observe that models combining MRI and genetic data achieved slightly higher performances than those using only MRI and that, among them the multilevel model achieved the best performance. Nevertheless, these differences are small and likely remain within the error margin. Also, the C-index was higher when combining MRI with all AD-related genes than when combining MRI and only the APOE gene but, again, the difference is small and likely within the error margin. The ROC(t) curves at different time points for our model and the

deep survival analysis model are shown on Figure 4.

G. Visualization of interactions between genes and brain regions for disease prediction

One interesting aspect of our approach is that one can look at the interactions between genes and brain regions, by examining the model parameters. Specifically, we displayed the model parameters as follows:

- on Figure 5, the matrix of interactions between genes and brain regions. Specifically, we displayed a reduced matrix presenting only one value per gene (instead of one value

per SNP), $\bar{\mathbf{W}} \in \mathcal{M}_{|I|,L}(\mathbb{R})$, where for each brain region j and gene ℓ , $\bar{\mathbf{W}}_{j,\ell} = \|\mathbf{W}_{s,G_\ell}\|_2$.

- (b) on Figure 6, selected rows of \mathbf{W} , displayed as Manhattan plots. For each selected brain region, the figure shows the SNPs that interact with this brain region to predict the disease.
- (c) on Figure 7(a), negative coefficients of the intercept $\beta_{\mathcal{I}}$. Each coefficient corresponds to one brain region.
- (d) on Figure 7(b), negative coefficients of a column of \mathbf{W} . Again, each coefficient corresponds to one brain region. Specifically, we show here the column of SNP rs429358, belonging to the gene *APOE*.

The matrix \mathbf{W} has a sparse structure, which was expected given the penalty used in the model. Among the different genes, *APOE* is the one which interacts with the largest number of regions. Nevertheless, one can observe that it does not interact with all regions. There are some regions with which it does not interact at all, in particular the different parts of the ventricular system. Moreover, its interaction with cortical regions is not uniform. Indeed, as can be seen on Fig. 7 (b), it interacts mainly with temporal and parietal, and less with primary motor, sensitive and visual areas. Some other genes (*TNF* and *TNFI*) interact with many brain regions, highlighting their potential importance in the progression of AD. On the other hand, the interactions of other genes are mostly restricted to a small set of brain regions: the entorhinal cortex, the hippocampus and the amygdala (which are part of the medial temporal lobe), the external temporal lobe and the ventricles. Note also that some genes do not interact with any brain region, thereby showing that they do not contribute much to the prediction of progression to AD.

IV. DISCUSSION

In this paper, we developed a novel method for disease prediction, that can integrate multimodal genetic and brain imaging data. Our approach is based on survival analysis, a technique which is ideally suited for predicting the time at which patients will develop the disease. We introduce an original multilevel approach that takes into account potential interactions between genes and brain regions, but also the structure of the different types of data through the use of specific penalties for each modality. Interestingly, the model was able to highlight the interactions between brain regions and genes in disease progression.

Various methods have been designed for disease prediction from medical imaging data. Many of these works have dealt with this question in the context of prediction of AD among patients with MCI (e.g. [44], [45], [46], [47], [48], [49], [38] and [2] for reviews on that topic). Most often, this is tackled as a classification problem: a date is fixed (e.g. 36 months) and one aims at classifying between patients who converted before that date and those who did not. This approach is interesting when using a fixed time frame is adequate (for example for a clinical trial with a fixed duration). Nevertheless, it may not be ideal for clinical applications because the date is arbitrary. For instance, in the above example, patients converting at 40 months would be considered non converters. One could

consider doing multiple classification tasks at different dates but the predictions may be inconsistent across the different dates.

Survival analysis is better suited to such tasks than classification because it allows to directly predict the date of conversion. It is also more adapted than standard regression because it can deal with censored data. Survival analysis is a classical statistical approach but its application in medical imaging has remained more limited. There are several applications of standard survival models, such as the Cox model, to different measurements including cognitive scores and MRI measures in order to predict conversion to AD among MCI patients [41], [40], [43], [50], [51], [53], [54]). Nevertheless, the standard Cox model is not adapted to handling high-dimensional data. Thung et al [52] have proposed a method for predicting time-to-conversion using low-rank affinity pursuit denoising and matrix completion, combining medical imaging (MRI, PET) and clinical scores.

We compared our method to various variations of the Cox model with different penalties and different types of input data. We also compared it to a recently proposed deep learning approach (deep survival analysis) [8]. The most striking result is that the use of MRI (either in isolation or when combined with genetics) resulted in a strong improvement when compared to genetic, age or gender (from 0.12 to 0.21 points of iAUC). Interestingly, age and gender alone were poor predictors of progression to AD. This may be counter-intuitive because age is well known as a major risk factor for AD. For instance, a comprehensive analysis based on the Rotterdam study showed that age alone was a good predictor of AD [10]. Nevertheless, it should be noted that this study is based on a general community population while ours includes patients with MCI at baseline. In our data, the baseline age of MCI who later progressed to AD is similar to that of stable MCI (74.5 vs 75.2 years). Therefore, it is not surprising that age was not a good predictor in our study. Predicting progression to AD among MCI patients or within a general population (whose majority of participants are healthy) are two different situations and thus we believe that our results are not contradictory with those reported in [10].

Compared to a standard Cox model (either using only MRI or combining MRI and genetics), we found that our approach resulted in a slightly superior performance. We also obtained slightly higher performances compared to the deep survival analysis [8]. However, the differences are small and are likely to be within the error margin. A similar observation can be made when comparing the combination of MRI with all AD-related genes or only with *APOE*. *APOE* is the genetic variant that is the most strongly associated to AD and has important discriminative power for detecting AD. However, in our experiments on the prediction of the conversion date in MCI patients, the performance of the model combining MRI and *APOE* was comparable to that of models using MRI alone. Thus, we cannot conclude that adding *APOE* allows improving prediction of the conversion date in MCI patients. More generally, our results do not allow concluding that the addition of genetic data results in higher predictive performance. Because we used a cross-validation setting, we could not perform

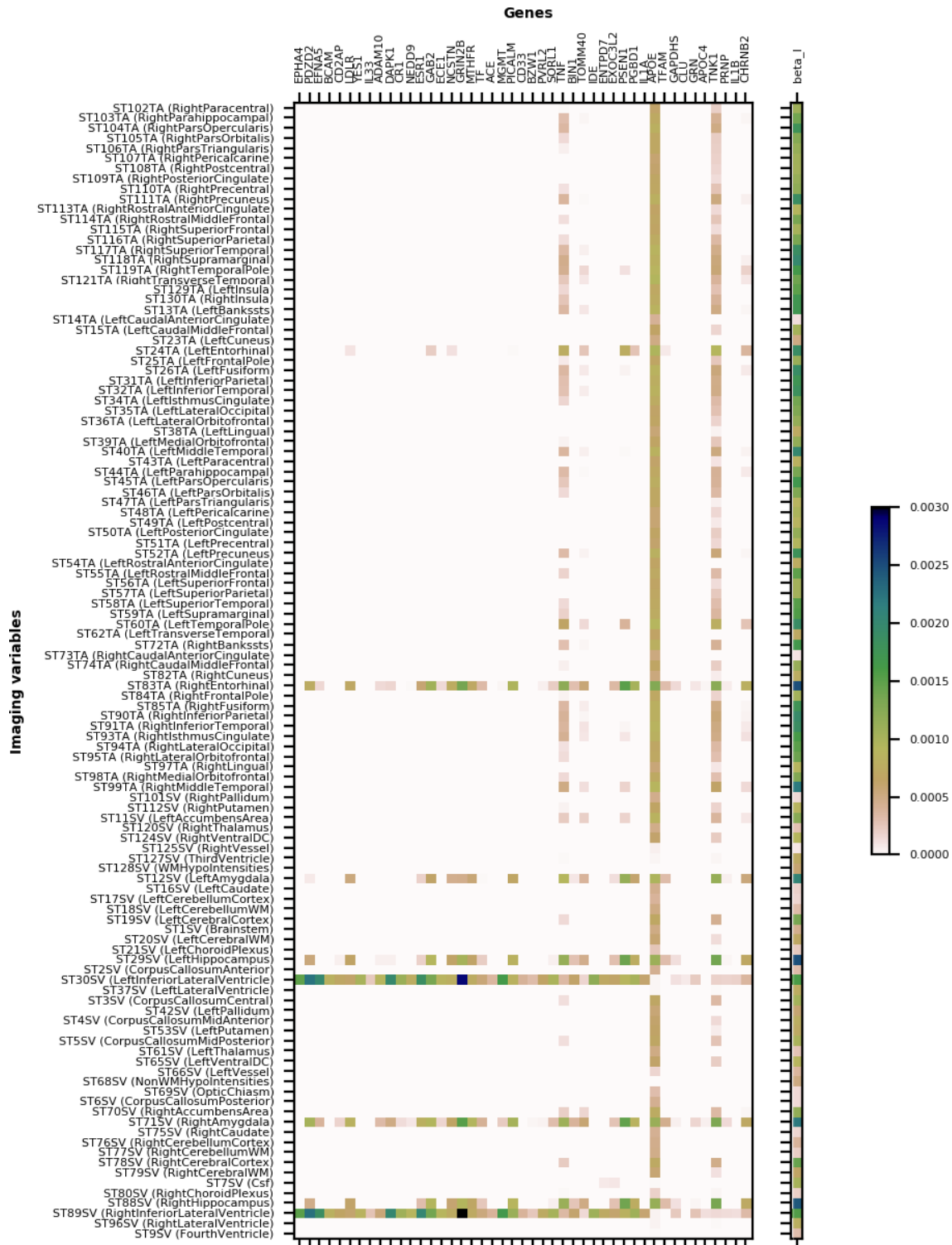


Fig. 5: Parameters learnt in the survival model. Reduced matrix \bar{W} of interactions between genes and brain regions

hypothesis testing to compare the performances. Indeed, there is no unbiased estimator of the variance of the performance in a cross-validation setting [64] and, as a result, the application of hypothesis testing can lead to unvalid inferences. Even though some alternative tests have been proposed [6], their validity is far from being universal and, in particular, they have not been proposed for survival models and metrics such

as those used in our paper. Performing hypothesis testing in a rigorous way would thus require an additional test set from a different cohort of MCI patients with imaging and genetics. This is left for future work. It is interesting to observe that our results are at least comparable to that of a recently proposed deep learning method for survival analysis [8]. One can note that the C-index achieved by deep survival analysis

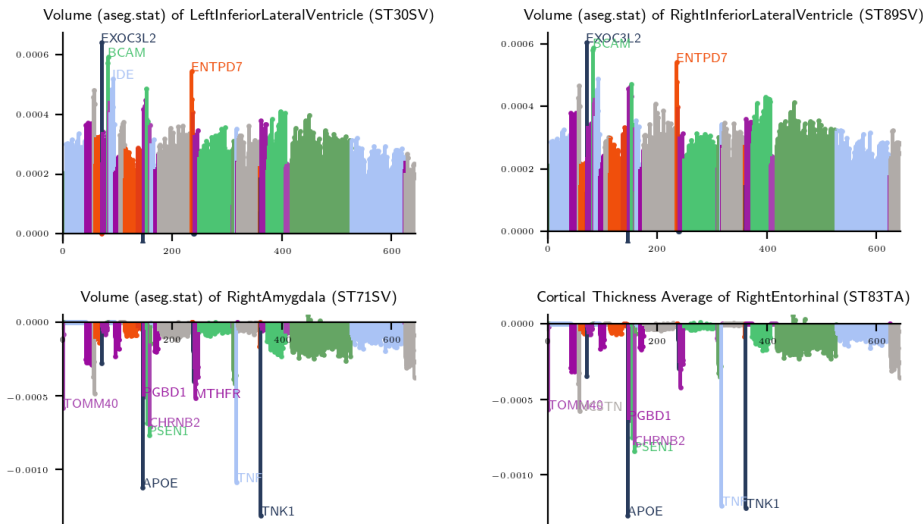


Fig. 6: Interaction with genes for a set of brain regions. The figure displays, as Manhattan plots, selected rows of \mathbf{W} . For each selected brain region, the figure shows the SNPs that interact with this brain region to predict the disease.

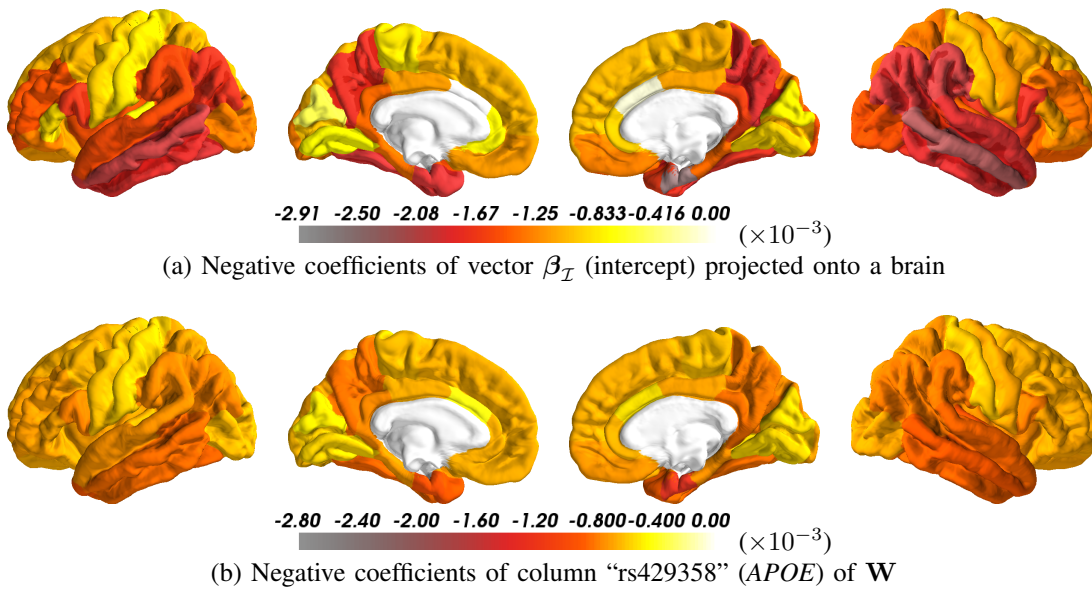


Fig. 7: (a) Negative coefficients of the intercept $\beta_{\mathcal{I}}$. Each coefficient corresponds to one cortical region. (b) Interactions between all cortical regions of SNP rs429358, belonging to the gene *APOE* (negative coefficients of the column of \mathbf{W} corresponding to SNP rs429358). Again, each coefficient corresponds to one brain region.

on our data was lower than that reported in their paper (they report a C-index of 0.72 when using only MCI patients and MRI data) [8]. There are several possible explanations to this difference. We believe that the most likely explanation is that they used a different population, in particular a much larger sample (1211 MCI patients vs 346 in our study) obtained by combining different datasets. Another possibility is that the difference comes from the features extracted from MRI data since they used voxel-based processing based on SPM and averaging into the regions of the AAL/BNA atlases while we used the Freesurfer processing and the Desikan-Killiany atlas, as provided by the ADNI. Investigating the effect of MRI

preprocessing on the results is out of the scope of our paper. The multilevel approach allowed to highlight interactions between imaging and genetic data. The idea for introducing a multilevel approach was to be able to model the case where some genes modulate the contribution of different imaging regions to the progression to AD. For instance, it could be that the pattern of regions which is predictive is different for patients who carry the *APOE* $\epsilon 4$ allele and for those who don't. If that would be the case, this would not be taken into account by an additive model. Larger coefficients in the interaction matrix mean that the interaction between a given brain region and a given SNP (or gene, when looking at the

reduced matrix) contributes to the prediction of the progression to AD. Unsurprisingly, the *APOE* gene, which is the main genetic risk factor for AD [58], interacted with many regions and in particular with temporal and parietal associative regions. This is highly consistent with the typical topographical pattern of atrophy and tau pathology of AD patients, as highlighted by many studies (e.g [55], [56], [57]). One intriguing result is that *APOE* did not interact with the ventricles. Ventricular volume is well-known to be associated with dementia (e.g. [59]) but is not specific of AD and is found in many other types of dementia including Parkinson's disease dementia [60], fronto-temporal dementia [61] or dementia with Lewy bodies [62]. In our results, the contribution of ventricles to prediction seems to be influenced by the *APOE* gene. The neurobiological implications of these results are beyond the scope of the present methodological paper.

We proposed a fast optimization procedure based on a proximal gradient. With 1,107 SNPs and 112 brain regions, it takes 8 minutes to train the model for fixed hyperparameters on a standard laptop (MacBook Pro, 2014 model). The computational time is linear in the number of SNPs. We indeed observed that the training took 16 minutes with 2107 and 41 minutes with 5107 SNPs. With all 600,000 SNPs, it would be around 80 hours. Note that these figures were obtained with a standard, relatively old, laptop. The cross-validation and hyperparameter search can be done in parallel thus the model using all SNPs is trainable within a reasonable time using a small-size computing cluster.

The limitations of our study are as follows. First, we did not consider death as a competing risk with Alzheimer's disease. This could be done using the framework of multistate survival models [63]. Nevertheless, one should not that death was not observed in the dataset used in our experiments. More generally, we could consider other competing risks such as other forms of dementia or even other diseases. Another limitation is that the diagnosis of AD in the ADNI database is based on clinical evaluations (McKhann criteria [42]) and does not include pathophysiological biomarkers. This can lead to misdiagnoses, the most likely misdiagnosis being that the underlying biological cause of the dementia is not AD. Nevertheless, we would like to point out that our focus is on predicting the future occurrence of dementia in patients with mild impairment at baseline. The underlying pathophysiological cause could be determined with amyloid and/or tau biomarkers, when available. Finally, it should be noted that, in general, coefficients from predictive models cannot be interpreted for inference as could be done for example in an association study. Indeed, it can be for instance that some interactions are associated with the outcome but are not included in the model because they would not increase the prediction accuracy or vice-versa. Therefore, the interaction coefficients of our model should be interpreted with caution. However, such limitation is true for any predictive model.

In conclusion, we proposed a new approach for disease prediction that can integrate imaging and genetic data and model their interaction. Its application to Alzheimer's disease showed that it can reveal interactions between genes and imaging variables in disease prediction. It is a generic technique

and can potentially be useful in other diseases.

ACKNOWLEDGMENT

The authors would like to thank Theodoros Evgeniou for many useful insights.

REFERENCES

- [1] D. Aschwanden *et al.* "Predicting Cognitive Impairment and Dementia: A Machine Learning Approach". *J Alzheimers Dis.* 2020;75(3):717-728.
- [2] M. Ansart *et al.* "Predicting the progression of mild cognitive impairment using machine learning: A systematic, quantitative and critical review", *Medical Image Analysis*, vol. 67, p. 101848, Jan. 2021.
- [3] K. Li and S. Luo, "Dynamic prediction of Alzheimer's disease progression using features of multiple longitudinal outcomes and time-to-event data", *Statistics in Medicine*, vol. 38, no. 24, pp. 4804–4818, Oct. 2019.
- [4] J. Lin, K. Li, and S. Luo, "Functional survival forests for multivariate longitudinal outcomes: Dynamic prediction of Alzheimer's disease progression", *Statistical Methods in Medical Research*, vol. 30, no. 1, pp. 99–111, Jan. 2021.
- [5] C. Platero and M. C. Tobar, "Longitudinal survival analysis and two-group comparison for predicting the progression of mild cognitive impairment to Alzheimer's disease," *Journal of Neuroscience Methods*, vol. 341, p. 108698, Jul. 2020.
- [6] C. Nadeau and Y. Bengio, "Inference for the Generalization Error", *Machine Learning*, volume 52, pp. 239–281, 2003
- [7] H. Zhou, J. Jiang, J. Lu, M. Wang, H. Zhang, C. Zuo, "Dual-Model Radiomic Biomarkers Predict Development of Mild Cognitive Impairment Progression to Alzheimer's Disease", *Frontiers in Neuroscience*, vol. 12, p. 1045, 2018.
- [8] T. Nakagawa *et al.* "Prediction of conversion to Alzheimer's disease using deep survival analysis of MRI images," *Brain Communications*, vol. 2, no. fcaa057, Jan. 2020.
- [9] D. Jarrett, J. Yoon, and M. van der Schaar, "Dynamic Prediction in Clinical Survival Analysis Using Temporal Convolutional Networks". *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 2, pp. 424–436, Feb. 2020.
- [10] S. Licher *et al.* "External validation of four dementia prediction models for use in the general community-dwelling population: a comparative analysis from the Rotterdam Study," *European Journal of Epidemiology*, vol. 33, no. 7, pp. 645–655, Jul. 2018.
- [11] P. Lu and O. Colliot. "Multilevel Modeling with Structured Penalties for Classification from Imaging Genetics data" in *3rd MICCAI Workshop on Imaging Genetics (MICGen 2017)*, Sep 2017, Quebec City, Canada. pp.230-240.
- [12] P. Lu and O. Colliot. "Multilevel Survival Analysis with Structured Penalties for Imaging Genetics data" in *SPIE Medical Imaging*, Feb 2020, Houston, USA.
- [13] R. C. Petersen *et al.* "Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization", *Neurology*, 2010 Jan 19; 74(3). pp.201–209.
- [14] T. Hastie, R. Tibshirani, and M. Wainwright. *Statistical Learning with Sparsity - The Lasso and Generalizations*, vol. 143. CRC Press, 2015.
- [15] Y. Ming and L. Yi. "Model selection and estimation in regression with grouped variables", *J. R. Statist. Soc. B*, part 1 **68**, 2006, pp. 49–67.
- [16] L. Meier, S. van de Geer, and P. Bühlmann. "The group lasso for logistic regression", *J. R. Statist. Soc. B* **70**, 2008, pp. 53–71.
- [17] L. Jacob, G. Obozinski, and J.-P. Vert. "Group lasso with overlap and graph lasso". In *Proceedings of the 26th International Conference on Machine Learning*, 2009.
- [18] A. Beck and M. Teboulle. "Gradient-based algorithms with applications to signal recovery problems". In Palomar, D.P., Eldar, Y.C. (eds.), *Convex Optimization in Signal Processing and Communications*, Cambridge University Press, Cambridge, 2010, pp. 42–88.
- [19] M. Gönen and E. Alpaydin. "Multiple kernel learning algorithms". *J. Mach. Learn. Res.* **12**, 2011, pp.2211–2268.
- [20] M. Kloft, U. Brefeld, S. Sonnenburg, and A. Zien. " ℓ_p -norm multiple kernel learning". *J. Mach. Learn. Res.* **12**, 2011, pp.953–997.
- [21] E.L. Kaplan and P. Meier. "Nonparametric Estimation from Incomplete Observations". *Journal of the American Statistical Association*, **53**, No. 282, Jun 1958, pp.457–481.
- [22] J. Liu, V.D. Calhoun. "A review of multivariate analyses in imaging genetics". *Front. Neuroinform.* **8**(29), 2014, pp. 1–11.

- [23] N.K. Batmanghelich, A. Dalca, G. Quon, M. Sabuncu, P. “Probabilistic modeling of imaging, genetics and diagnosis”. *IEEE TMI* **35**, 2016, pp.1765–1779.
- [24] O. Kohannim *et al.* “Discovery and replication of gene influences on brain structure using LASSO regression”. *Front. Neurosci.* **6**(115), 2012.
- [25] M. Silver *et al.* “Fast identification of biological pathways associated with a quantitative trait using group lasso with overlaps”. *Stat. Appl. Genet. Mol. Biol.* **11**(1), 2012, pp.1–40.
- [26] M. Silver, E. Janousova, X. Hua, P.M. Thompson, G. Montana, and ADNI. “Identification of gene pathways implicated in Alzheimer’s disease using longitudinal imaging phenotypes with sparse regression”. *NeuroImage* **63**(3), 2012, pp.1681–1694.
- [27] H. Wang, F. Nie, H. Huang, S.L. Risacher, A.J. Saykin, and L. Shen. “Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning”. *Bioinformatics* **28**(12), 2012, pp.127–136.
- [28] J. Peng, L. An, X. Zhu, Y. Jin and D. Shen. “Structured sparse kernel learning for imaging genetics based Alzheimer’s Disease Diagnosis”. In: Ourselein, S. *et al.* (eds.) *MICCAI 2016*. LNCS, vol. 9901, Springer, Cham, 2016, pp. 70–78.
- [29] L. Brand *et al.* “Joint High-Order Multi-Task Feature Learning to Predict the Progression of Alzheimer’s Disease”. In: Frangi A. *et al.* (eds) *MICCAI 2018*. LNCS, vol 11070, Springer, Cham, 2018.
- [30] F. Aiolli and M. Donini. “EasyMKL: a scalable multiple kernel learning algorithm”. *Neurocomputing*, **169**, 2015, pp.215–224.
- [31] M. Lorenzi *et al.* “Partial least squares modelling for imaging-genetics in Alzheimer’s Disease: plausibility and generalization”. In *IEEE ISBI*, 2016.
- [32] L. Du *et al.* “A novel structure-aware sparse learning algorithm for brain imaging genetics”. In: Golland, P. *et al.* (eds.), *MICCAI 2014*, LNCS, vol. 8675, 2014, Springer, Cham, pp.329–336.
- [33] L.E. Chambless *et al.* “Estimation of time-dependent area under the ROC curve for long-term risk prediction”. *Statistics in Medicine*, 2006.
- [34] E. Graf. “Assessment and Comparison of Prognostic classification schemes for survival data”. *Statistics in Medicine*, **18**, 1999, pp.2529–2545.
- [35] H. Steck *et al.* “On Ranking in Survival Analysis: Bounds on the Concordance Index”. In *NIPS* **20**, 2007.
- [36] V. Razvan *et al.*, “TADPOLE Challenge: Prediction of Longitudinal Evolution in Alzheimer’s Disease”, arXiv:1805.03909, 2018.
- [37] A.S. Panayides, M. Pattichis, S. Leandrou, C. Pitris, A. Constantinidou, and C.S. Pattichis. “Radiogenomics for Precision Medicine With A Big Data Analytics Perspective”. *IEEE Journal of Biomedical and Health Informatics*, 2018. pp. 329–336.
- [38] S. “A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer’s disease and its prodromal stages”. *NeuroImage*, **155**, 2017, pp.530–548.
- [39] H.M. Bovelstad “Survival prediction from clinico-genomic models - a comparative study”. *BMC Bioinformatics*, **10**(413), 2009.
- [40] E. D. Anderson *et al.* “Cognitive variability: A marker for incident MCI and AD: An analysis for the Alzheimer’s Disease Neuroimaging Initiative”. *Alzheimer’s & Dementia: Diagnosis, Assessment & Disease Monitoring*, **4**, 2016, pp.47–55.
- [41] K. Li, R. O’Brien, M. Lutz, S. Luo, and ADNI. “A prognostic model of Alzheimer’s disease relying on multiple longitudinal measures and time-to-event data”. *Alzheimer’s & Dementia*, 2017.
- [42] G. McKhann G *et al.* “Clinical diagnosis of Alzheimer’s disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer’s Disease”. *Neurology*. **34** (7): 939–44.
- [43] K. Liu, K. Chen, L. Yao, and X. Guo. “Prediction of Mild Cognitive Impairment Conversion Using a Combination of Independent Component Analysis and the Cox Model”. *Frontiers in Human Neuroscience*, **11**:33, 2017.
- [44] 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, pp.1415–1422.
- [45] C. Plant *et al.* “Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer’s disease”. *NeuroImage*, **50**, 2010, pp.162–174.
- [46] M. Liu, D. Zhang, and D. Shen. “View-centralized multi-atlas classification for Alzheimer’s disease diagnosis”. *Hum. Brain Mapp.* **36**, 2015, pp.1847–1865.
- [47] C. Davatzikos, P. Bhatt, L.M. Shaw, K.N. Batmanghelich, and J.Q. Trojanowski. “Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification”. *Neurobiol. Aging* **32** (2322), 2011, pp.2319–2327.
- [48] E. Moradi, A. Pepe, C. Gaser, H. Huttunen, and J. Tohka. “Machine learning framework for early MRI-based Alzheimer’s conversion prediction in MCI subjects”. *NeuroImage*, **104**, 2015, pp.398–412.
- [49] B. Cheng, M. Liu, D. Zhang, B.C. Munsell, and D. Shen. “Domain transfer learning for MCI conversion prediction”. *IEEE Trans. Biomed. Eng.* **62**, 2015, pp.1805–1817.
- [50] C. Platero, M.E. López, M.D. Carmen Tobar, M. Yus, F. Maestu. “Discriminating Alzheimer’s disease progression using a new hippocampal marker from T1-weighted MRI: The local surface roughness”. *Hum Brain Mapp.*, 2019, **40**(5), pp.1666–1676.
- [51] K. Kauppi *et al.* “Combining Polygenic Hazard Score With Volumetric MRI and Cognitive Measures Improves Prediction of Progression From Mild Cognitive Impairment to Alzheimer’s Disease”. *Front Neurosci.* **12**:260, 2018.
- [52] K.H. Thung, P.T. Yap, E. Adeli, S.W. Lee, D. Shen, and ADNI. “Conversion and time-to-conversion predictions of mild cognitive impairment using low-rank affinity pursuit denoising and matrix completion”. *Med Image Anal.*, 2018, **45**, pp.68–82.
- [53] K. Brueggen *et al.* “Basal Forebrain and Hippocampus as Predictors of Conversion to Alzheimer’s Disease in Patients with Mild Cognitive Impairment - A Multicenter DTI and Volumetry Study”. *J Alzheimers Dis.*, **48**(1), 2015, pp.197–204.
- [54] M. Ewers *et al.* “Prediction of conversion from mild cognitive impairment to Alzheimer’s disease dementia based upon biomarkers and neuropsychological test performance”. *Neurobiol Aging*, **33**(7):1203–14, 2012.
- [55] J.C. Baron *et al.* “In Vivo Mapping of Gray Matter Loss With Voxel-Based Morphometry in Mild Alzheimer’s Disease”. *NeuroImage*, **14**(2), 2001, pp.298–309.
- [56] A. Delacourte *et al.* “The Biochemical Pathway of Neurofibrillary Degeneration in Aging and Alzheimer’s Disease”. *Neurology*. **52**(6), 1999, pp.1158–65.
- [57] A.T. Du *et al.* “Different regional patterns of cortical thinning in Alzheimer’s disease and frontotemporal dementia”. *Brain*, **130**(Pt 4), 2007, pp.1159–66.
- [58] J.C. Lambert *et al.* “Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease”. *Nature Genetics*, **45**(12), 2013, pp. 1452–1458.
- [59] O.T. Carmichael *et al.* “Ventricular volume and dementia progression in the Cardiovascular Health Study”, *Neurobiol Aging* 2007 Mar;**28**(3):389–97.
- [60] R. Camicioli *et al.* “Ventricular dilatation and brain atrophy in patients with Parkinson’s disease with incipient dementia”, *Mov Disord* 2011 Jul;**26**(8):1443–50.
- [61] T.P. Tavares *et al.* “Ventricular volume expansion in presymptomatic genetic frontotemporal dementia”, *Neurology* 2019 Oct **29**;93(18):e1699–e1706.
- [62] R. Barber, C. Ballard, I.G. McKeith, A. Gholkar, and J.T. O’Brien. “MRI volumetric study of dementia with Lewy bodies: A comparison with AD and vascular dementia”. In *Neurology*, vol 54, issue 6, 2000, pp.1304–1309
- [63] K. Leffondré, C. Touraine, C. Helmer, and P. Joly. “Interval-censored time-to-event and competing risk with death: is the illness-death model more accurate than the Cox model?” In *International Journal of Epidemiology*, **42**(4), pp.1177–1186
- [64] J. Bengio and Y. Grandvalet. “No unbiased estimator of the variance of k-fold cross-validation”. In *Journal of machine learning research*, 2004.
- [65] J.H. Lanskey *et al.* “Can neuroimaging predict dementia in Parkinson’s disease?”, In *Brain*, **141**(9), 2018, pp.2545–2560.
- [66] R. Iniesta *et al.* “Antidepressant drug-specific prediction of depression treatment outcomes from genetic and clinical variables”, *Sci Rep.*, 2018, **8**(1):5530.
- [67] S. Khanna *et al.* “Using Multi-Scale Genetic, Neuroimaging and Clinical Data for Predicting Alzheimer’s Disease and Reconstruction of Relevant Biological Mechanisms.” *Sci Rep.*, 2018, **8**(11173).
- [68] A. N. Kamarudin, T. Cox and R. Kolumunage-Dona “Time-dependent ROC curve analysis in medical research: current methods and applications.” *BMC medical research methodology*, 2017, **17**(1):53.