A Wide and Deep Neural Network for Survival Analysis from Anatomical Shape and Tabular Clinical Data

Sebastian Pölsterl, Ignacio Sarasua, Benjamín Gutiérrez-Becker, and Christian Wachinger

Artificial Intelligence in Medical Imaging (AI-Med),
Department of Child and Adolescent Psychiatry,
Ludwig-Maximilians-Universität, Munich, Germany
sebastian.poelsterl@med.uni-muenchen.de

Abstract. We introduce a wide and deep neural network for prediction of progression from patients with mild cognitive impairment to Alzheimer's disease. Information from anatomical shape and tabular clinical data (demographics, biomarkers) are fused in a single neural network. The network is invariant to shape transformations and avoids the need to identify point correspondences between shapes. To account for right censored time-to-event data, i.e., when it is only known that a patient did not develop Alzheimer's disease up to a particular time point, we employ a loss commonly used in survival analysis. Our network is trained end-to-end to combine information from a patient's hippocampus shape and clinical biomarkers. Our experiments on data from the Alzheimer's Disease Neuroimaging Initiative demonstrate that our proposed model is able to learn a shape descriptor that augments clinical biomarkers and outperforms a deep neural network on shape alone and a linear model on common clinical biomarkers.

1 Introduction

Alzheimer's disease (AD) is a neurodegenarative disorder and the most common form of dementia diagnosed in people over 65 years of age. Initially, patients suffer from short memory loss, until progressive deterioration eventually requires patients to be completely dependent upon caregivers due to severe impairment of cognitive and motor abilities [38,45,1]. Mild cognitive impairment (MCI) is a pre-dementia stage which is characterized by clinically significant cognitive decline, but without impairing daily live [41,29]. Although subjects with MCI are at an increased risk of developing dementia due to AD, a significant portion of patients with MCI remain stable and do not progress [41]. The pathophysiological processes of this transition are complex and not fully understood, but previous studies showed that changes in certain biomarkers precede the onset of cognitive symptoms by many years [25]. Important biomarkers include brain atrophy measured by magnetic resonance images (MRI), levels of cortical amyloid deposition obtained from cerebrospinal fluid (CSF), and glucose uptake of neurons measured

by fluorodeoxyglucose positron emission tomography (FDG-PET) (see [44] for a detailed overview). To stop or slow down the progression to dementia, it is vital to identify those patients that are at an increased risk for rapid progression from MCI to AD. In particular, several previous studies have established strong morphological changes in the hippocampus associated to the progression of dementia [18,19,51,50,20].

We study progression to Alzheimer's disease by explicitly modelling the timing of this transition and by considering the finite follow-up time and drop-out of patients in clinical studies using techniques from survival analysis (also called time-to-event analysis). Survival analysis differs from traditional machine learning in the fact that parts of the training data can only be partially observed – they are *censored*. If a patient withdraws from the study, is lost to follow-up, or did not develop AD during the study period, the patient's time of progression is *right censored*, i.e., it is unknown whether the patient has or has not progressed after the study ended. Only if a patient develops AD during the study period, one can record the exact time of this event – it is *uncensored*.

In this paper, we propose for the first time a wide and deep neural network for survival analysis that learns to identify patients at high risk of progressing to AD by fusing information from 3D hippocampus shape and tabular clinical data. To the best of our knowledge, no one has previously attempted to learn a deep survival model on 3D anatomical shape representations in an end-to-end fashion. In our experiments on data from the Alzheimer's Disease Neuroimaging Initiative, we demonstrate by fusing information we can more accurately predict AD converters than a baseline deep network on shapes and a Cox's proportional hazards model on clinical data.

2   Related Work

Most previous work formulates progression analysis from MCI to AD as a classification problem within a fixed time horizon such as 3 years (see e.g. [4,9,11,40,48]). The major downside of this approach is that such a model cannot generalize to other time spans, and that censored conversion times are ignored during training. Instead, it is statistically more appropriate to explicitly incorporate censored event times using methods from survival analysis. Several authors used survival analysis techniques by combining information from various modalities such as structural MRI, FDG-PET, genetics, and neuropsychological tests [3,12,13,14,15,27,31,34,46,49,51,53]. All of these approaches compute features from high-dimensional imaging data in a pre-processing step, before training a linear survival model. They differ with respect to the type and extend of computed features, which range from volume measurements of a few brain regions [15] to voxel-based analysis [49]. In addition, we note that extensive prior work aims to identify healthy controls, patients with MCI, and patients with AD by casting it as a three-way classification problem and using multi-view machine learning techniques; we refer interested readers to the review in [36].

In contrast, this work focuses on multi-view learning to predict progression from MCI to AD, which has been formulated as a classification problem within a fixed time period in [35,47,52,54]. [52] propose to use sparsity-inducing penalties to combine features extracted from MRI and PET images with CSF measurements and neuropsychological tests. MCI to AD conversion within 2 years was studied in [35]. They propose to learn from features extracted from MRI and FDG-PET, and CSF measurements by view-aligned hypergraph learning. The approach in [47] uses stability-weighted low-rank matrix completion to impute missing values in MRI and PET features, and neuropsychological tests. They consider right censored conversion times as missing values and try to impute the actual (unobserved) time of conversion via matrix completion. In [54], the authors propose a missing-data-aware approach to learn from MRI, PET, and genetics by learning a common and multiple modality-specific latent feature representations. To the best or our knowledge, the only previous work that employed multi-view learning for survival analysis was presented in [42] for predicting adverse events in cancer and heart disease.

Using neural networks for survival analysis originated in the late 1990s in the work of [2,5,16,33], who studied relatively simple networks with one hidden layer applied to tabular data. The first deep survival model was proposed in [26] and builds on the loss proposed in [16]. The only previous work that investigated deep learning for MCI to AD conversion from multi-modal data is [30,37]. Both approaches consider a classification problem within a fixed time frame, which ignores censoring of conversion times. In addition, the features in [30] were pre-computed from MRI and not learned end-to-end. In [37], a deep network is proposed that learns from 3D patches of MRI and FDG-PET at multiple scales.

Finally, [20] proposed a deep neural network operating on point clouds of multiple neuroanatomical shapes. They study diagnosis of MCI and AD patients rather than progression, and do not consider demographics or clinical biomarkers in their model.

3   Methods

We present a wide and deep neural network for learning from right censored time-to-event data (see fig. 1). Our model takes a point cloud representation of an anatomical shape and tabular data as input. The deep part of the network is a PointNet [43] that learns features describing the 3D geometric structure of the left hippocampus. The wide part of the network takes demographics and clinical biomarkers and their interactions. The network is trained to fuse both types of information in and end-to-end fashion using a survival analysis loss appropriate for right censored event times. First, we are going to describe PointNet, which constitutes the deep part of the network, before showing how it can be integrated with tabular clinical data for survival analysis.

3.1   Learning from Anatomical Shape

We represent anatomical shapes as point clouds that represent a 3D geometric structure as a set of coordinates. Point clouds avoid the combinatorial irregularities and complexities of meshes, and thus are easier to learn from. However, the network needs to be constructed in a way to consider that a point cloud is just an unordered set of points that is invariant to permutations of its members. To this end, we employ PointNet [43], which is illustrated in fig. 1 and described in more detail below.

The i-th point cloud $\mathcal{P}_i$ is represented by a set of K 3D coordinates $\mathcal{P}_i = \{\mathbf{p}_{i_1}, \ldots, \mathbf{p}_{i_K}\}$ with $\mathbf{p}_{i_k} \in \mathbb{R}^3$ being the x, y, and z coordinates. To be invariant to permutations of the input set, the symmetric max pooling operator across all embedding vectors of points is used. We first pass each individual coordinate vector through a multilayer perceptron $\text{MLP}_{\text{point}}$ with shared weights among all points, thus projecting each 3D point to a higher dimensional representation. These representations are aggregated using the max pooling operator across all points, which ensures that our downstream survival analysis task is invariant to permutation:

$$\text{POINTNET}(\mathcal{P}_i) = \text{MAXPOOL}\left(\text{MLP}_{\text{points}}(\mathbf{p}_{i_1}), \ldots, \text{MLP}_{\text{points}}(\mathbf{p}_{i_K})\right). \quad (1)$$

$\text{MLP}_{\text{point}}$ is a three-layer network with 64, 128, and 400 dimensional outputs, respectively, with rectified linear units (ReLU) and batch normalization [23]. Hence, we extract 400 features that globally describe the input anatomical shape.

In order to make our network invariant to rotation of the input point cloud, we use an affine transformation network that outputs a rotation matrix $\mathbf{T} \in \mathbb{R}^{3 \times 3}$ which is multiplied by the raw 3D coordinates of input points. This transformation is learned in a data-dependent manner by using an additional POINTNET network that learns to predict the optimal $\mathbf{T}$ for each individual point cloud. The global feature vector computed by POINTNET is fed to three fully-connected layers with 200, 100, and 9 units, ReLU activation function and batch normalization, respectively. Finally, we modify the vanilla PointNet in (3) by transforming individual points by the output of the transformation network:

$$\text{TRANSFORM}(\mathcal{P}_i) = \text{MAXPOOL}\left(\text{MLP}_{\text{points}}(\mathbf{p}_{i_1}), \ldots, \text{MLP}_{\text{points}}(\mathbf{p}_{i_K})\right),$$
$$\boldsymbol{\varphi}_{i_k} = \text{TRANSFORM}(\mathcal{P}_i)\mathbf{p}_{i_k}, \quad (2)$$
$$\text{POINTNET}(\mathcal{P}_i) = \text{MAXPOOL}\left(\text{MLP}_{\text{points}}(\boldsymbol{\varphi}_{i_1}), \ldots, \text{MLP}_{\text{points}}(\boldsymbol{\varphi}_{i_K})\right).$$

3.2   Wide and Deep Neural Network

After obtaining a global latent representation of an anatomical shape, we can further learn high-level descriptors of point clouds by feeding the output of the max pooling operation to an MLP. In addition, we can leverage routine clinical patient information to predict progression to Alzheimer's disease. Typically, such information consists of feature vectors that are either dense (e.g. biomarker concentrations), or sparse (e.g. one-hot encoded genetic alterations). Compared
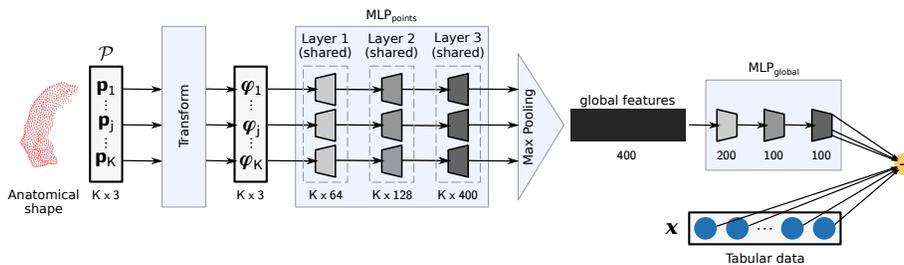
Fig. 1. Wide and Deep PointNet Architecture. The network takes a point cloud representation $\mathcal{P}$ of the left hippocampus with K points, applies a transformation, and then aggregates point features by max pooling. The global feature vector is processed by a global MLP outputting a 100-dimensional latent representation that is fused with tabular clinical data using a linear model.

to individual points in a point cloud, clinical information already contains rich information for which we do not need to learn a highly abstract latent representation. In fact, most clinical research relies on linear models, which allow for easy interpretation of individual feature's contribution to the overall prediction.

Here, we jointly train a linear model on clinical information with a deep PointNet on anatomical shapes using a wide and deep architecture [8]. While the deep component learns a complex latent representation of anatomical shape, the linear component models known clinical variables $\mathbf{x} \in \mathbb{R}^d$ associated with Alzheimer's disease. In particular, we can easily incorporate gene-gene (epistasis) and gene–environment interactions by using a cross-product transformation $\phi(\mathbf{x})$ [8]. Thus, the final patient-level latent representation is given by

$$\mu(\mathbf{x}_i, \mathcal{P}_i) = \mathbf{w}_{\text{wide}}^\top \text{CONCAT}\left(\mathbf{x}_i, \phi(\mathbf{x}_i)\right)$$
$$+ \mathbf{w}_{\text{deep}}^\top \text{MLP}_{\text{global}}(\text{POINTNET}(\mathcal{P}_i)), \quad (3)$$

where CONCAT denotes vector concatenation, POINTNET is the global feature vector from (2), $\text{MLP}_{\text{global}}$ is a three-layer MLP with 200, 100, and 100 units, ReLU activation and batch normalization, and $\mathbf{w}_{\text{wide}}$ and $\mathbf{w}_{\text{deep}}$ are weights to be learned.

3.3 Survival Analysis

Our overall objective is to predict progression from mild cognitive impairment to Alzheimer's disease from right censored time-to-event data, which demands for proper training algorithms that take this unique characteristic into account. More formally, we denote by $t_i > 0$ the time of an event (Alzheimer's disease), and $c_i > 0$ the time of censoring of the i-th patient. Due to right censoring, it is only possible to observe $y_i = \min(t_i, c_i)$ and $\delta_i = I(t_i \le c_i)$ for every patient, with $I(\cdot)$ being the indicator function and $c_i = \infty$ for uncensored records. Hence, training our survival model is based on a dataset comprising quadruplets $(\mathcal{P}_i, \mathbf{x}_i, y_i, \delta_i)$

for $i = 1, \ldots, n$. After training, the survival model ought to predict a risk score of experiencing an event based on a point cloud and a set of clinical features. As loss function, we employ the loss proposed in [16], which is an extension of Cox's proportional hazards model [10] to neural networks. Let $\boldsymbol{\Theta}$ denote the set of all parameters of the wide and deep neural network (3), then we want to solve

$$\underset{\boldsymbol{\Theta}}{\arg\min} \quad \sum_{i=1}^{n} \delta_i \left[\mu(\boldsymbol{x}_i, \mathcal{P}_i \,|\, \boldsymbol{\Theta}) - \log \left(\sum_{j \in \mathcal{R}_i} \exp(\mu(\boldsymbol{x}_j, \mathcal{P}_i \,|\, \boldsymbol{\Theta})) \right) \right], \quad (4)$$

where $\mathcal{R}_i = \{j \,|\, y_j \geq t_i\}$ denotes the risk set, i.e., the set of patients who were still free of Alzheimer's disease shortly before time point t_i.

4   Experiments

4.1   Data

In our experiments, we are using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [24]. ADNI was launched in 2003 as a public-private partnership with the primary goal to test whether longitudinal MRI and PET imaging combined with other biomarkers, clinical and neuropsychological assessments to measure the progression of MCI and early AD. For up-to-date information, see www.adni-info.org. We selected 397 subjects with MCI at baseline and at least one follow-up visit. Magnetic resonance images of all subjects were processed with FreeSurfer [17] to obtain segmentations, which were subsequently pre-processed using the grooming operations included in ShapeWorks [7] to obtain smooth hippocampi surfaces. We used left hippocampus shapes represented as point clouds comprised of 1024 points. For tabular clinical data, we used age, gender, education, CSF, FDG-PET, and AV45-PET. CSF measurements included levels of beta amyloid 42 peptides ($A\beta_{42}$), total tau protein (T-tau), and Tau phosphorylated at threonine 181 (p-Tau$_{181}$). We augment age to account for non-linear effects by using a natural B-spline expansion with four degrees of freedom and an interaction term between age and gender [22]. Education, which is a categorical variable, was encoded using orthogonal polynomial coding. In addition, we considered left hippocampus volume (normalized by intra-cranial volume) as estimated by FreeSurfer [17] from MRI scans of the brain.

4.2   Model Training

We trained our deep and wide network using Adam [28] for 120 epochs with weight decay. We tuned hyper-parameters (size of PointNet's global feature vector, size of $\mathbf{w}_{\text{deep}}$, weight decay, learning rate schedule, β_1 of Adam) using Bayesian black-box optimization by computing the model's performance on the validation set [32]. Data is randomly split into three parts: 80% for training, 10% for validation, and 10% for testing. We repeated this process 10 times with different splits. The performance of all methods was estimated by Harrell's concordance index (c

index), which is identical to the area under the receiver operating characteristics curve if the outcome is binary and no censoring is present [21]. As baseline model, we selected a linear Cox's proportional hazards model (CoxPH) [10] trained on tabular clinical data. The baseline model was trained once on tabular clinical data only (see above), and once with the volume of left hippocampus included as additional feature. We note that CoxPH and our model optimize the same loss during training. Therefore, differences in performance stem from the ability of our model to directly incorporate 3D anatomical shape information.
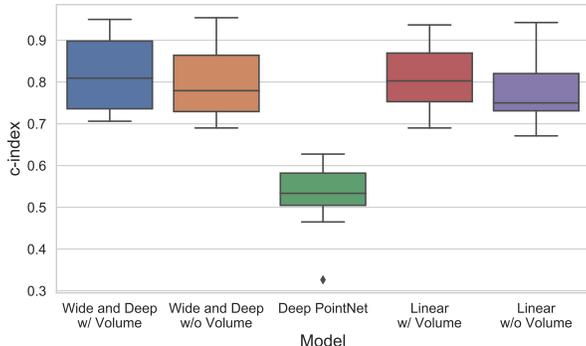
5   Results

Fig. 2. Performance of individual models across ten random splits of the data. w/ Volume: tabular data includes left hippocampus volume. w/o Volume: tabular data does *not* include left hippocampus volume.

The performance of our deep and wide network and baseline models is summarized in fig. 2. It shows that tabular clinical makers with a median c index of 0.750 are already strong predictors of conversion from MCI to AD. When including hippocampus volume as additional feature, the median c index increased to 0.803. Using a deep PointNet solely using hippocampus shape and ignoring any clinical variables resulted in a c index of 0.534. Our deep and wide network achieved a median c index of 0.780 without hippocampus volume, and 0.809 with hippocampus volume. The latter is the model with highest median c index and outperforms the linear model with hippocampus volume on 6 of 10 splits. This shows that when jointly learning a deep PointNet, it is able to learn a powerful global descriptor of hippocampus shape that augments clinical features for MCI-to-AD progression. Moreover, our results confirm that hippocampus volume is a useful independent predictor that cannot be fully captured by anatomical shape alone, as described previously [50].

We can also compare the coefficients of the linear models with the linear part of our wide and deep neural network. The coefficients can be directly interpreted
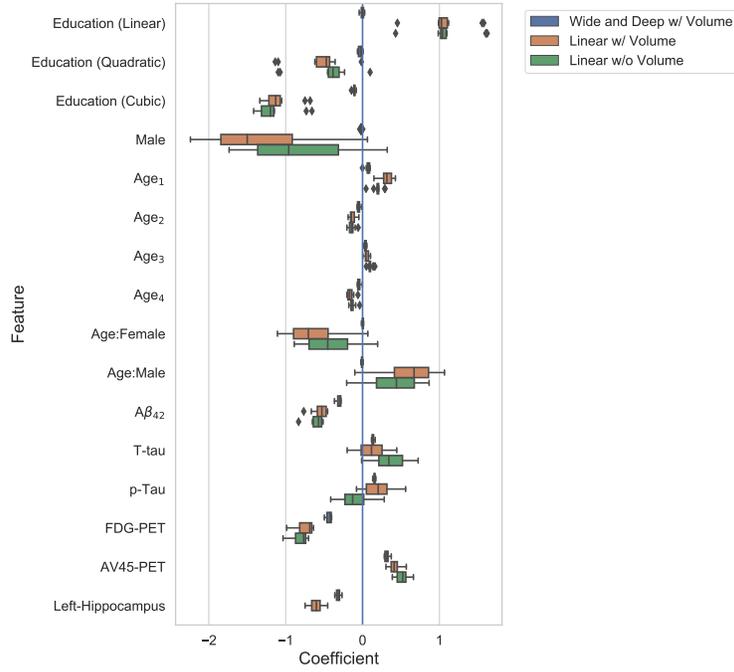
Fig. 3. Comparison of coefficients associated with tabular clinical features. Additional eight orthogonal polynomial encodings of education have been omitted from this plot. w/ Volume: tabular data includes left hippocampus volume. w/o Volume: tabular data does *not* include left hippocampus volume.

in terms of log-hazard ratio, which is a measure of effect a variable has on survival, similar to log-odds ratio in logistic regression. The coefficients across all folds are depicted in fig. 3. All models agree with respect to which features are contributing to increased/decreased hazard of AD, as indicated by the coefficients' sign, except for p-Tau. The linear model without hippocampus volume associated higher p-Tau levels with a decrease in hazard (on average) compared to the other models, which is surprising because hyperphosphorylation of tau is a marker for AD [6]. The most important clinical features (in terms of magnitude) are gender and education for both linear models, but have only minor importance for the deep and wide network. Similar behavior can be observed for age-gender interactions. In addition, increased hippocampus volume has a relatively high importance and is associated with a decreased hazard of AD. It is ranked third for the deep and wide network and eleven for the linear model. FDG-PET has the biggest effect for the wide and deep network and is also among the top 4 features for the linear models. From a clinical perspective, this result is reassuring as reduction of metabolic activity in cortical regions has been associated with AD [39]. Finally, we note that the variability of coefficients across splits is smaller

for the deep and wide neural network compared to the linear model. We believe this is an effect of using weight decay during optimization, which penalizes large coefficients.

6   Conclusion

We proposed a wide and deep neural network that fuses 3D anatomical shape and tabular clinical variables for the prediction of MCI-to-AD conversion. We trained a model end-to-end using a survival loss that properly accounts for right censored time of conversion. Our experiments demonstrate that the proposed architecture is able to learn a global shape descriptor that augments clinical variables and leads to improved prediction performance.

References

1. Albert, M.S., DeKosky, S.T., Dickson, D., Dubois, B., Feldman, H.H., Fox, N.C., Gamst, A., Holtzman, D.M., et al.: The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & dementia : the journal of the Alzheimer's Association 7, 270–279 (2011)
2. Bakker, B., Heskes, T.: A neural-Bayesian approach to survival analysis. In: 9th Int. Conf. Artif. Neural Networks (ICANN). pp. 832–837 (1999)
3. Barnes, D.E., Cenzer, I.S., Yaffe, K., Ritchie, C.S., Lee, S.J.: A point-based tool to predict conversion from mild cognitive impairment to probable Alzheimer's disease. Alzheimer's & Dementia 10(6), 646–655 (2014)
4. Beheshti, I., Demirel, H., Matsuda, H., Alzheimer's Disease Neuroimaging Initiative: Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm. Computers in biology and medicine 83, 109–119 (2017)
5. Biganzoli, E., Boracchi, P., Mariani, L., Marubini, E.: Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach. Stat. Med. 17(10), 1169–1186 (1998)
6. Blennow, K., Vanmechelen, E., Hampel, H.: CSF Total tau, $A\beta 42$ and Phosphorylated tau Protein as Biomarkers for Alzheimer's Disease. Molecular Neurobiology 24(1-3), 087–098 (2001)
7. Cates, J., Fletcher, P.T., Styner, M., Hazlett, H.C., Whitaker, R.: Particle-based shape analysis of multi-object complexes. In: Medical image computing and computer-assisted intervention (MICCAI). pp. 477–485 (2008)

8. Cheng, H.T., Ispir, M., Anil, R., Haque, Z., Hong, L., Jain, V., Liu, X., Shah, H., et al.: Wide & Deep Learning for Recommender Systems. In: Proc. of the 1st Workshop on Deep Learning for Recommender Systems (DLRS) (2016)

9. Chételat, G., Landeau, B., Eustache, F., Mézenge, F., Viader, F., de la Sayette, V., Desgranges, B., Baron, J.C.: Using voxel-based morphometry to map the structural changes associated with rapid conversion in MCI: a longitudinal MRI study. NeuroImage 27, 934–946 (2005)

10. Cox, D.R.: Regression models and life tables (with discussion). Journal of the Royal Statistical Society. Series B (Statistical Methodology) 34, 187–220 (1972)

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

12. Da, X., Toledo, J.B., Zee, J., Wolk, D.A., Xie, S.X., Ou, Y., Shacklett, A., Parmpi, P., et al.: Integration and relative value of biomarkers for prediction of MCI to AD progression: spatial patterns of brain atrophy, cognitive scores, APOE genotype and CSF biomarkers. NeuroImage. Clinical 4, 164–173 (2014)

13. Desikan, R.S., Cabral, H.J., Fischl, B., Guttmann, C.R.G., Blacker, D., Hyman, B.T., Albert, M.S., Killiany, R.J.: Temporoparietal MR imaging measures of atrophy in subjects with mild cognitive impairment that predict subsequent diagnosis of Alzheimer disease. American journal of neuroradiology 30, 532–538 (2009)

14. Desikan, R.S., Cabral, H.J., Settecase, F., Hess, C.P., Dillon, W.P., Glastonbury, C.M., Weiner, M.W., Schmansky, N.J., et al.: Automated MRI measures predict progression to Alzheimer's disease. Neurobiology of aging 31, 1364–1374 (2010)

15. Devanand, D.P., Pradhaban, G., Liu, X., Khandji, A., Santi, S.D., Segal, S., Rusinek, H., Pelton, G.H., et al.: Hippocampal and entorhinal atrophy in mild cognitive impairment: Prediction of Alzheimer disease. Neurology 68(11), 828–836 (2007)

16. Faraggi, D., Simon, R.: A neural network model for survival data. Stat. Med. 14(1), 73–82 (1995)

17. Fischl, B.: FreeSurfer. NeuroImage 62(2), 774–781 (2012)

18. Frisoni, G.B., Ganzola, R., Canu, E., Rub, U., Pizzini, F.B., Alessandrini, F., Zoccatelli, G., Beltramello, A., Caltagirone, C., Thompson, P.M.: Mapping local hippocampal changes in Alzheimer's disease and normal ageing with MRI at 3 Tesla. Brain 131(12), 3266–3276 (2008)

19. Gerardin, E., Chételat, G., Chupin, M., Cuingnet, R., Desgranges, B., Kim, H.S., Niethammer, M., Dubois, B., Lehéricy, S., Garnero, L., Eustache, F., Colliot, O., Initiative, A.D.N.: Multidimensional classification of hippocampal shape features discriminates alzheimer's disease and mild cognitive impairment from normal aging. NeuroImage 47, 1476–1486 (2009)

20. Gutiérrez-Becker, B., Wachinger, C.: Deep multi-structural shape analysis: application to neuroanatomy. In: Medical image computing and computer-assisted intervention (MICCAI). pp. 523–531 (2018)

21. Harrell, F.E., Califf, R.M., Pryor, D.B., Lee, K.L., Rosati, R.A.: Evaluating the yield of medical tests. Journal of the American Medical Association 247, 2543–2546 (1982)

22. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer, second edn. (2009)

23. Ioffe, S., Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: Proc. of the 32nd International Conference on Machine Learning. pp. 448–456 (2015)

24. Jack, C.R., Bernstein, M.A., Fox, N.C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P.J., et al.: The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. Journal of Magnetic Resonance Imaging 27(4), 685–691 (2008)
25. Jack, C.R., Knopman, D.S., Jagust, W.J., Petersen, R.C., Weiner, M.W., Aisen, P.S., Shaw, L.M., Vemuri, P., et al.: Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. The Lancet Neurology 12(2), 207–216 (2013)
26. Katzman, J.L., Shaham, U., Bates, J., Cloninger, A., Jiang, T., Kluger, Y., Bates, J., Jiang, T., Kluger, Y.: DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. BMC Med. Res. Methodol. 18(24) (2018)
27. Kauppi, K., Fan, C.C., McEvoy, L.K., Holland, D., Tan, C.H., Chen, C.H., Andreassen, O.A., Desikan, R.S., et al.: Combining Polygenic Hazard Score With Volumetric MRI and Cognitive Measures Improves Prediction of Progression From Mild Cognitive Impairment to Alzheimer's Disease. Frontiers in Neuroscience 12 (2018)
28. Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization. In: 3rd Int. Conf. on Learning Representations (ICLR) (2015)
29. Langa, K.M., Levine, D.A.: The diagnosis and management of mild cognitive impairment: a clinical review. JAMA 312, 2551–2561 (2014)
30. Lee, G., Nho, K., Kang, B., Sohn, K.A., Kim, D., Alzheimer's Disease Neuroimaging Initiative: Predicting Alzheimer's disease progression using multi-modal deep learning approach. Scientific reports 9, 1952 (2019)
31. Li, K., O'Brien, R., Lutz, M., Luo, S., Alzheimer's Disease Neuroimaging Initiative: A prognostic model of Alzheimer's disease relying on multiple longitudinal measures and time-to-event data. Alzheimer's & dementia : the journal of the Alzheimer's Association 14, 644–651 (2018)
32. Liaw, R., Liang, E., Nishihara, R., Moritz, P., Gonzalez, J.E., Stoica, I.: Tune: A Research Platform for Distributed Model Selection and Training (2018)
33. Liestøl, K., Andersen, P.K., Andersen, U.: Survival analysis and neural nets. Stat. Med. 13(12), 1189–1200 (1994)
34. Liu, K., Chen, K., Yao, L., Guo, X.: Prediction of Mild Cognitive Impairment Conversion Using a Combination of Independent Component Analysis and the Cox Model. Frontiers in human neuroscience 11, 33 (2017)
35. Liu, M., Zhang, J., Yap, P.T., Shen, D.: View-aligned hypergraph learning for Alzheimer's disease diagnosis with incomplete multi-modality data. Medical image analysis 36, 123–134 (2017)
36. Liu, X., Chen, K., Wu, T., Weidman, D., Lure, F., Li, J.: Use of multimodality imaging and artificial intelligence for diagnosis and prognosis of early stages of Alzheimer's disease. Translational research : the journal of laboratory and clinical medicine 194, 56–67 (2018)
37. Lu, D., Popuri, K., Ding, G.W., Balachandar, R., Beg, M.F., Alzheimer's Disease Neuroimaging Initiative: Multimodal and Multiscale Deep Neural Networks for the Early Diagnosis of Alzheimer's Disease using structural MR and FDG-PET images. Scientific reports 8, 5697 (2018)
38. McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack, C.R., Kawas, C.H., Klunk, W.E., Koroshetz, W.J., et al.: The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease.

Alzheimer's & dementia : the journal of the Alzheimer's Association 7(3), 263–9 (2011)

39. Minoshima, S., Giordani, B., Berent, S., Frey, K.A., Foster, N.L., Kuhl, D.E.: Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease. Annals of Neurology 42(1), 85–94 (1997)

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

41. Petersen, R.C.: Mild Cognitive Impairment. New England Journal of Medicine 364(23), 2227–2234 (2011)

42. Pölsterl, S., Conjeti, S., Navab, N., Katouzian, A.: Survival analysis for high-dimensional, heterogeneous medical data: Exploring feature extraction as an alternative to feature selection. Artificial intelligence in medicine 72, 1–11 (2016)

43. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 652–660 (2017)

44. Scheltens, P., Blennow, K., Breteler, M.M.B., de Strooper, B., Frisoni, G.B., Salloway, S., Van der Flier, W.M.: Alzheimer's disease. The Lancet 388(10043), 505–517 (2016)

45. Sperling, R.A., Aisen, P.S., Beckett, L.A., Bennett, D.A., Craft, S., Fagan, A.M., Iwatsubo, T., Jack, C.R., et al.: Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimer's & dementia : the journal of the Alzheimer's Association 7(3), 280–92 (2011)

46. Teipel, S.J., Kurth, J., Krause, B., Grothe, M.J.: The relative importance of imaging markers for the prediction of Alzheimer's disease dementia in mild cognitive impairment — Beyond classical regression. NeuroImage: Clinical 8, 583–593 (2015)

47. Thung, K.H., Adeli, E., Yap, P.T., Shen, D.: Stability-Weighted Matrix Completion of Incomplete Multi-modal Data for Disease Diagnosis. Medical image computing and computer-assisted intervention (MICCAI) pp. 88–96 (2016)

48. Tong, T., Gao, Q., Guerrero, R., Ledig, C., Chen, L., Rueckert, D., Alzheimer's Disease Neuroimaging Initiative: A Novel Grading Biomarker for the Prediction of Conversion From Mild Cognitive Impairment to Alzheimer's Disease. IEEE transactions on bio-medical engineering 64, 155–165 (2017)

49. Vemuri, P., Weigand, S.D., Knopman, D.S., Kantarci, K., Boeve, B.F., Petersen, R.C., Jack, C.R.: Time-to-event voxel-based techniques to assess regional atrophy associated with MCI risk of progression to AD. NeuroImage 54, 985–991 (2011)

50. Wachinger, C., Reuter, M., Initiative, A.D.N., et al.: Domain adaptation for Alzheimer's disease diagnostics. Neuroimage 139, 470–479 (2016)

51. Wachinger, C., Salat, D.H., Weiner, M., Reuter, M., Initiative, A.D.N.: Whole-brain analysis reveals increased neuroanatomical asymmetries in dementia for hippocampus and amygdala. Brain 139(12), 3253–3266 (2016)

52. Zhang, D., Shen, D., Alzheimer's Disease Neuroimaging Initiative: Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease. NeuroImage 59, 895–907 (2012)

53. Zhou, H., Jiang, J., Lu, J., Wang, M., Zhang, H., and, C.Z.: Dual-Model Radiomic Biomarkers Predict Development of Mild Cognitive Impairment Progression to Alzheimer's Disease. Frontiers in Neuroscience 12 (2019)

54. Zhou, T., Liu, M., Thung, K.H., Shen, D.: Latent Representation Learning for Alzheimer's Disease Diagnosis with Incomplete Multi-modality Neuroimaging and Genetic Data. IEEE Transactions on Medical Imaging pp. 1–1 (2019)