

Deep learning predictions of survival based on MRI in amyotrophic lateral sclerosis



Hannelore K. van der Burgh^a, Ruben Schmidt^a, Henk-Jan Westeneng^a, Marcel A. de Reus^b, Leonard H. van den Berg^{a,1}, Martijn P. van den Heuvel^{b,*,1}

^aDepartment of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Heidelberglaan 100, PO Box 85500, 3508 GA, Utrecht, Netherlands

^bDepartment of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Heidelberglaan 100, PO Box 85500, 3508 GA, Utrecht, Netherlands

ARTICLE INFO

Article history:

Received 28 July 2016

Received in revised form 28 September 2016

Accepted 10 October 2016

Available online 11 October 2016

Keywords:

Deep learning

Neural network

Amyotrophic lateral sclerosis

White matter connectivity

Survival

Prediction

ABSTRACT

Amyotrophic lateral sclerosis (ALS) is a progressive neuromuscular disease, with large variation in survival between patients. Currently, it remains rather difficult to predict survival based on clinical parameters alone. Here, we set out to use clinical characteristics in combination with MRI data to predict survival of ALS patients using deep learning, a machine learning technique highly effective in a broad range of big-data analyses. A group of 135 ALS patients was included from whom high-resolution diffusion-weighted and T1-weighted images were acquired at the first visit to the outpatient clinic. Next, each of the patients was monitored carefully and survival time to death was recorded. Patients were labeled as short, medium or long survivors, based on their recorded time to death as measured from the time of disease onset. In the deep learning procedure, the total group of 135 patients was split into a training set for deep learning ($n = 83$ patients), a validation set ($n = 20$) and an independent evaluation set ($n = 32$) to evaluate the performance of the obtained deep learning networks. Deep learning based on clinical characteristics predicted survival category correctly in 68.8% of the cases. Deep learning based on MRI predicted 62.5% correctly using structural connectivity and 62.5% using brain morphology data. Notably, when we combined the three sources of information, deep learning prediction accuracy increased to 84.4%. Taken together, our findings show the added value of MRI with respect to predicting survival in ALS, demonstrating the advantage of deep learning in disease prognostication.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Amyotrophic lateral sclerosis (ALS) is a progressive neuromuscular disease, heterogeneous in terms of symptom development, disease onset and disease progression (Chiò et al., 2011; Ravits and La Spada, 2009). ALS patients display, on average, a survival time of 3–4 years after onset of symptoms (del Aguila et al., 2003; Hardiman et al., 2011). To date, clinical characteristics such as site of onset, respiratory status, ALS Functional Rating Scale (ALSFRS) scores (Cedarbaum et al., 1999) and C9orf72 phenotype status (i.e. a disease-causing repeat expansion mutation in ALS (DeJesus-Hernandez et al., 2011; Renton et al., 2011)) are shown to have some predictive power for prediction of survival (Chiò et al., 2009; Elamin et al., 2015; Scotton et al., 2012; Wolf et al., 2014, 2015). Prognosis based on these markers, however, often remains too uncertain to be implemented in clinical practice (Elamin et al., 2015) as motor neuron loss might already occur before clinical weakness can be measured (Simon et al., 2014). This stresses the importance of the development of new (objective)

markers and neuroimaging techniques might provide such markers and improve prognostication (Turner et al., 2011). Reliable prediction of survival at the first clinical MRI appointment would provide highly valuable information for patients and care providers.

Neuroimaging data has been used to separate ALS patients from healthy controls. In these approaches, machine learning techniques (e.g. support vector machines, (Cristianini and Shawe-Taylor, 2000)) have been used to group patients and controls on the basis of changes in motor and extra-motor resting-state functional connectivity, reaching an overall accuracy of 87% (Fekete et al., 2013). Also other resting-state networks have been used in machine learning approaches with an accuracy of 72% (Welsh et al., 2013). In ALS, cortical thinning and subcortical changes have been related to disease progression, putting forward these changes as a biomarker of ALS (Agosta et al., 2012; Mezzapesa et al., 2013; Turner and Verstraete, 2015; Verstraete et al., 2012; Walhout et al., 2015; Westeneng et al., 2015). For this reason, cortical thickness has been used as an alternative imaging metric to separate patients from controls, reaching accuracies ranging from 60 to 75% (Ahmed et al., 2015; Folland-Ross et al., 2015; Greenstein et al., 2012; Lerch et al., 2008). Notwithstanding the importance of exploring classification approaches to distinguish between patients and controls,

* Corresponding author.

E-mail address: m.p.vandenheuvel@umcutrecht.nl (M.P. van den Heuvel).

¹ Authors contributed equally.

predicting *disease course*, i.e. going beyond the establishment of patient-control identification, is presumably a more difficult problem, and arguably clinically more relevant. In the present study we therefore set out to explore the use of clinical characteristics in combination with MRI-based metrics of connectivity and brain morphology. White matter brain connectivity was derived from diffusion-weighted connectome imaging and brain morphology (i.e. cortical thickness and subcortical volume), was extracted from T1-weighted images. We used deep learning, a powerful technique shown to be of great value in many classification problems (Dean et al., 2012; Hinton et al., 2012; Krizhevsky et al., 2012; Mohamed et al., 2012; Wu et al., 2015), with Google's visual search algorithm and Google's AlphaGo program as well-known examples of applications of deep learning (Dean et al., 2012; Krizhevsky et al., 2012; Silver et al., 2016). Deep learning methods have shown high value in the fields of image classification (Krizhevsky et al., 2012; Wu et al., 2015), speech recognition (Hinton et al., 2012; Mohamed et al., 2012) as well as in elucidating complex relationships in MRI data (Plis et al., 2014).

Focusing on testing the predictive power of deep learning in differentiating between three survival duration subgroups (i.e. short, medium and long survivors), we assess the prediction accuracy of four deep learning networks which are based on 1. clinical data, 2. structural connectivity MRI data, 3. morphology MRI data and 4. a combined approach, in which clinical and imaging data are combined using layered deep learning. We show that the prediction accuracy of future survival time of ALS patients can be as high as 84% on the basis of combined clinical and neuroimaging data.

2. Materials and methods

2.1. Patients

A total dataset of 135 patients with sporadic ALS was included in this study (Table 1). Patients were diagnosed according to the El Escorial

criteria (Brooks et al., 2009) and recruited from the outpatient clinic for motor neuron diseases of the University Medical Center Utrecht. Parts of this dataset have been described in earlier publications in the context of examining patient-control group effects (Schmidt et al., 2014; Verstraete et al., 2011) (demographics given in Table 1). The included set involved data of patients that were either deceased ($n = 122$) or still alive with a disease duration of over 50 months ($n = 13$) at the time of analysis, providing the opportunity to test survival predictions using information from the first clinical MRI appointment.

The included clinical characteristics (Table 1) consisted of eight metrics. These comprised 1. site of disease onset, 2. age at disease onset and 3. time to diagnosis (i.e. the time from disease onset until the diagnosis of ALS was given). In addition, 4. the ALSFRS slope was included to provide an indication of disease progression based on the revised ALS Functional Rating Scale (ALSFRS-R) (Cedarbaum et al., 1999) and time T (in months) between symptom onset and first examination: slope = $(48 - \text{ALSFRS-R}) / T$ (Kimura et al., 2006; Kollewé et al., 2008; Qureshi et al., 2006). Other clinical variables taken into account were 5. the forced vital capacity (FVC), 6. C9orf72 phenotype status, 7. frontotemporal dementia (FTD) status as derived from the (revised) Neary criteria (Neary et al., 1998; Rascovsky et al., 2011) and 8. El Escorial criteria diagnostic category.

Among the group of 135 patients, there was no history of brain injury, psychiatric illness, epilepsy, or neurodegenerative diseases other than ALS. The Ethical Committee for human research of the University Medical Center Utrecht approved the study protocols and informed written consent according to the Declaration of Helsinki was obtained from each patient.

2.2. Short, medium, long survivors

Each of the 135 patients was categorized according to the true survival time (i.e. time between disease onset and death): *short* survivors

Table 1

Demographic and clinical characteristics of all study participants. Total dataset is divided in a training set, validation set and evaluation set.

	Total n = 135	Training n = 83	Validation n = 20	Evaluation n = 32	p-value ^a
Survival class distribution [n (%)]					0.994
Short	52 (38.5)	31 (37.3)	8 (40.0)	13 (40.6)	
Medium	52 (38.5)	32 (38.6)	8 (40.0)	12 (37.5)	
Long	31 (23.0)	20 (24.1)	4 (20.0)	7 (21.9)	
Age at first MRI (years)	61.7 (30.3–78.9)	61.3 (35.8–78.5)	63.0 (30.3–71.7)	62.2(46–78.9)	0.741
Sex (m/f)	99 / 36	60 / 23	16 / 4	23 / 9	0.765
Disease duration (months) ^b	21.8 (3.2–174.4)	20.1 (3.2–174.4)	24.0 (3.7–109)	25.0 (4–126.5)	0.976
Site of onset [n (%)]					0.164
Bulbar	32 (23.7)	24 (28.9%)	4 (20.0)	4 (12.5)	
Non-bulbar	103 (76.3)	59 (71.1)	16 (80.0)	28 (87.5)	
Age at onset (years)	59.9 (25.8–78.1)	59.6 (35.2–77.4)	61.0 (25.8–70.7)	60.1 (43.8–78.1)	0.852
TTD (months)	15.1 (1.3–149.6)	14.2 (1.3–149.6)	12.9 (2.7–75.9)	18.5 (4.0–120.0)	0.537
ALSFRS slope ^c	0.74 (0.05–4.08)	0.69 (0.06–3.70)	1.03 (0.12–4.08)	0.68 (0.05–3.49)	0.126
FVC (in %)	94.3 (59–142)	95.1 (59–142)	93.5 (62–121)	92.9 (65–124)	0.813
C9orf72 status [n (%)]					0.256
Long	8 (5.9)	3 (3.6)	1 (5.0)	4 (12.5)	
Wild type	117 (86.7)	73 (88.0)	19 (95.0)	25 (78.1)	
Unknown	10 (7.4)	7 (8.4)	0 (0.0)	3 (9.4)	
FTD status [n (%)]					0.082
ALS with FTD	3 (2.2)	1 (1.2)	0 (0.0)	2 (6.3)	
ALS without FTD	115 (85.2)	71 (85.5)	15 (75.0)	29 (90.6)	
Unknown	17 (12.6)	11 (13.3)	5 (25.0)	1 (3.1)	
El Escorial category [n (%)]					0.351
Definite ALS	21 (15.6)	15 (18.1)	1 (5.0)	5 (15.6)	
Not definite ALS	114 (84.4)	68 (81.9)	19 (95.0)	27 (84.4)	

Values are in mean (min-max) unless otherwise specified.

SD, standard deviation; TTD, time to diagnosis; FTD, frontotemporal dementia; FVC, forced vital capacity; ALSFRS-R, revised ALS functional rating scale.

^a p-values were computed using a one-way ANOVA (age at first MRI, disease duration, age at onset, TTD, ALSFRS slope, FVC) or chi-squared test (class distribution, sex, site of onset, C9orf72 phenotype status, FTD status, El Escorial category) among the four subsets.

^b Disease duration was measured from disease onset until date of first MRI scan.

^c ALSFRS slope was calculated using the formula $(48 - \text{ALSFRS-R score}) / \text{time between symptom onset and first examination (in months)}$.

with survival up to 25 months after disease onset, *medium* survivors with survival between 25 and 50 months after disease onset, and *long* survivors living over 50 months after disease onset (Elamin et al., 2015). The group of long survivors consisted of patients who either died after a disease duration of at least 50 months or were still alive and had a disease duration of at least 50 months at time of analysis.

2.3. Image acquisition and preprocessing

T1 and diffusion-weighted scans were acquired from all patients using a 3 Tesla Philips Achieva Medical Scanner with a SENSE receiver head-coil, described in detail by (Verstraete et al., 2011) and (Schmidt et al., 2014). A high-resolution T1-weighted image was acquired for anatomical reference by a 3D fast field echo using parallel imaging (TR/TE = 10/4.6 ms, flip-angle 8°, slice orientation: sagittal, voxel size = 0.80 × 0.75 × 0.75 mm, field of view = 176 × 240 × 240 mm covering the whole brain). For each subject, two sets of 30 weighted diffusion scans and 5 unweighted B0 scans were acquired with opposite *k*-space readouts (Andersson et al., 2003) using the following settings: parallel imaging SENSE p-reduction 3, high angular gradient set of 30 different weighted directions, TR/TE = 7035/68 ms, 2 × 2 × 2 mm voxel size, 75 slices, b = 1000 s/mm². Anatomical T1-weighted images were parcellated using FreeSurfer (V5.1.0) according to the Desikan-Killiany atlas (Desikan et al., 2006) dividing the segmented gray matter into 83 distinct brain regions (68 cortical regions (34 for each hemisphere), 14 subcortical areas, and the brainstem). Of the 68 cortical regions, cortical thickness was measured by computing the distances between the gray/white matter boundary and pial surface at each point on the cortical mantle (Fischl and Dale, 2000). Volumes of the 14 subcortical areas and the brainstem were computed with FreeSurfer's automated procedure for volumetric measurements (Fischl et al., 2002). Preprocessing of diffusion-weighted images included corrections for susceptibility and eddy-current distortions (Andersson and Skare, 2002; Andersson et al., 2003). Next, a tensor was fitted to the diffusion signals in each voxel and diffusion tensor imaging metrics, such as fractional anisotropy (FA) (Alexander et al., 2007), were derived. White matter tracts were reconstructed using Fiber Assignment by Continuous Tracking (FACT) (Mori et al., 1999; Mori et al., 2002; Mori and van Zijl, 2002); tracking was initiated by 8 seeds per white matter voxel and stopped using conditions as detailed in previous work (Schmidt et al., 2014). For each subject, an individual brain network was reconstructed by selecting the interconnecting tracts from the total cloud of reconstructed streamlines for each pair of regions included in the used cortical atlas (van den Heuvel and Sporns, 2011; Verstraete et al., 2011). We focused on white matter connectivity strength measured in terms of FA (from here on referred to as connection weight). In support of using this metric as a marker for disease effects, previous studies have extensively shown FA changes in ALS patients (Agosta et al., 2011; Ciccarelli et al., 2006; Menke et al., 2012; Schmidt et al., 2014; Senda et al., 2011; Turner et al., 2011; Verstraete et al., 2010; Verstraete et al., 2014). Moreover, the extent to which FA alterations are observed has been suggested to reflect distinct sequential ALS disease stages (Kassubek et al., 2014; Müller et al., 2016; Schmidt et al., 2015), and has been noted to mirror the pattern of phosphorylated 43 kDa TAR DNA-binding protein (pTDP-43, (Brettschneider et al., 2013)) aggregation. FA values of tracts interconnecting brain regions were stored in a weighted connectivity matrix.

2.4. Training, validation and test set

The total set of 135 datasets (i.e. 135 patients) was randomly divided into a training set (n = 83), a validation set (n = 20), and an independent evaluation set (n = 32), by first randomly selecting an evaluation set of the total dataset, and next by dividing the remaining subset of the data into a training set and a validation set, with the proportional splits between 70/30 and 80/20 (Crowther and Cox, 2005; Shahin et al.,

2004). The training, validation and evaluation sets had similar survival class distributions (Fig. 1) and were not significantly different for each of the eight clinical characteristics (Table 1).

2.5. Deep learning neural network

A deep learning approach on the basis of an artificial neural network was applied. In brief (see for details on the applied deep learning procedures below), the procedure included deep learning on the training set, with the validation set used to stop the training process on time to prevent overfitting of the classifier to the training set (Duda et al., 2001). The evaluation set was then used to assess the final performance of the trained neural network on an independent sample.

For clarity, we note that the term 'neural network' here is taken from the field of supervised learning (Bishop, 1995; Duda et al., 2001; Hinton et al., 2012; Larochelle et al., 2009) and does not refer to the concept of brain network as used in the field of connectomics (Bullmore and Sporns, 2009; van den Heuvel et al., 2012; van den Heuvel et al., 2008). *Deep learning networks* refer to the subclass of neural networks comprising multiple hidden layers (Hinton et al., 2012) that allow for a detailed input-to-output mapping. Due to the inclusion of multiple hidden layers, deep learning networks are particularly useful for modeling high-level abstractions from data (Larochelle et al., 2009).

In total, four deep learning networks were constructed (Fig. 2). These were based on 1. clinical data, 2. structural connectivity MRI data, 3. morphology MRI data and 4. a combination of the previous three information sources. The input vectors or features of these networks included the normalized clinical characteristics (8 in total), normalized cortical thickness and subcortical volumes derived from T1-weighted MRI and/or the connection weights as stored in the connectivity matrices. Normalization was performed using a min-max feature scaling in order to accelerate training (Priddy and Keller, 2005). Average values of features were imputed for missing values (i.e. either unknown clinical characteristics or connections that could not be detected in patients), with missing values being accounted for in an additional binary input vector. The output vectors, i.e. the classes that the deep learning

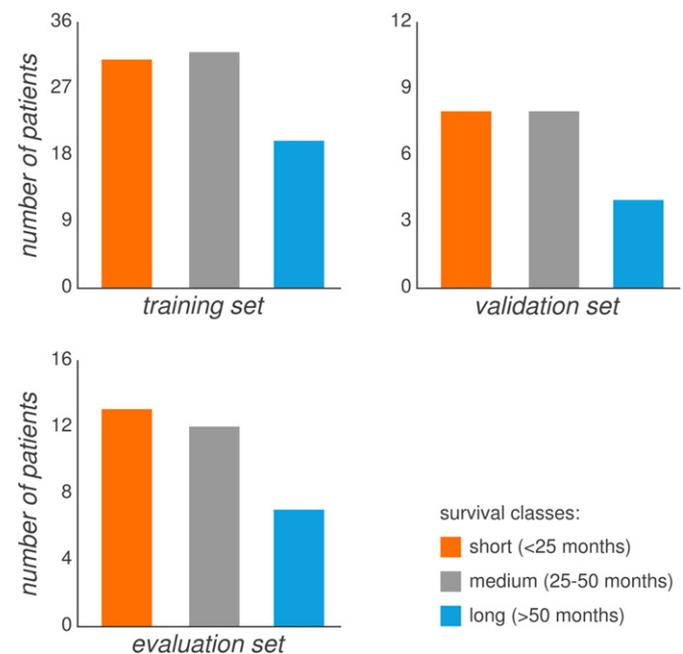


Fig. 1. Survival class distributions of the 135 subjects across the training, validation, and evaluation sets. The three survival categories (short, medium and long) are displayed for each data subset. The training set contained 83 patients and the validation set 20 patients. The evaluation set consisted of 32 patients. There are no significant differences in known survival class distributions and demographics between the three datasets (all $p > 0.05$, see Table 1 for more details).

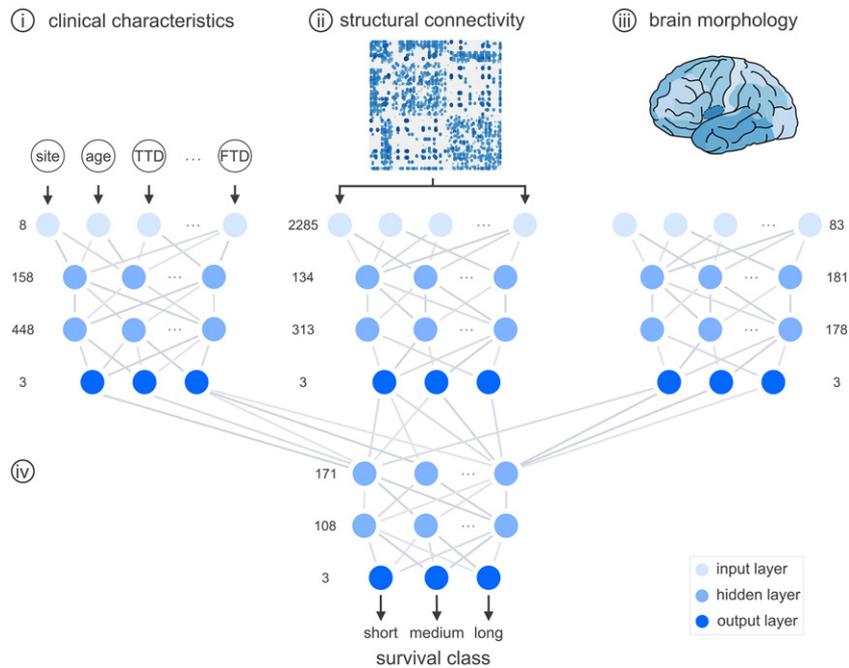


Fig. 2. Visualization of the prediction process using deep learning networks. (i) First, a network based on clinical characteristics (i.e. site of onset, age at onset, time to diagnosis, ALSFRS slope, FVC, C9orf72 phenotype status, FTD status and El Escorial criteria diagnostic category) is fitted, consisting of 158 and 448 hidden nodes. (ii) Next, a network is fitted based on connectivity matrices, storing white matter connectivity strengths in terms of FA, composing of 134 and 313 hidden nodes. (iii) Third, a network based on brain morphology (68 cortical thickness values and 15 subcortical volumes) is fitted, consisting of 181 and 178 hidden nodes. (iv) The output nodes of these three networks are used as input nodes for the combination network, consisting of 171 and 108 hidden nodes. From the output nodes of each network a class label can be derived, revealing the survival prediction of a patient.

network had to classify, represented the predefined classes of short, medium and long survivors. In what follows, we will describe the construction of the four deep learning networks, their training and their evaluation.

2.6. Deep learning network construction

1. Clinical deep learning

First, a deep learning network was constructed for the clinical characteristics (Fig. 2). The input vector of this network represented the eight clinical characteristics, including site of onset, age at onset, time to diagnosis, ALSFRS slope, FVC, C9orf72 phenotype status, FTD status and El Escorial criteria diagnostic category. Each layer in the deep learning network consisted of nodes and was connected to other layers using weighted edges (Duda et al., 2001). The number of input nodes was based on the number of clinical characteristics and the number of output nodes was based on the number of classes (short, medium, long). The number of hidden layers was set to two, in order to balance between the benefit of discovering more complex relations, the risk of overfitting, and training time and complexity (Karsoliya, 2012). The number of hidden nodes was set by means of a fine neuron grid search during the training phase (described below), with the number of hidden nodes in both layers varying from 1 to 500.

2. Structural connectivity deep learning

Second, a deep learning network was constructed for the structural connectivity MRI metric (Fig. 2). This second network employed connection weight of each reconstructed connection as an input node for the deep learning (2285 features in total: $83 \text{ brain regions} \times 82 / 2 = 3403$ possible connections of which 2285 are existing connections). The number of output nodes was set to three survival classes. Two hidden layers were used and the sizes of these layers were found using a fine neuron grid search in the same search domain as similarly set for clinical deep learning.

3. Morphology deep learning

Next, a deep learning network was constructed based on the cortical

thickness and subcortical volume measurements (Fig. 2). The input vector of the morphology network included 68 cortical thickness values and volume values of 14 subcortical regions and the brainstem, resulting in a total of 83 input nodes. The number of output nodes was set to three survival classes and two hidden layers were used. The size of these layers was set by means of a fine neuron grid search in the same search domain as described for structural connectivity deep learning.

4. Clinical-MRI combined deep learning

In addition to the clinical and MRI deep learning networks, a fourth network was constructed that combined the three information sources. The input layer of this network included each of the three output nodes of the clinical and MRI deep learning networks, thus consisted of nine input nodes (Fig. 2). The number of hidden nodes was found using a neuron grid search in the same range as for the other networks during the training phase. Output nodes (three nodes) represented survival class.

2.7. Network training

All four deep learning networks were trained using the following procedures. Each hidden node was assigned a non-linear (logistic) activation function (Hjelm et al., 2014). Feeding each training data point through the network produced the output vector of weights. This output vector was compared to the target values, with any difference in the predicted outcome and the real outcome (i.e. short, medium, long survivors) defined as error using the cross-entropy error function (Murphy, 2012; Rubinstein and Kroese, 2004). After presenting all examples of the training set to the network, the network weights were updated by backpropagation learning (Rumelhart et al., 1986) using the scaled conjugate gradient algorithm (Møller, 1993) to correct the weights in a direction that reduced the error of the network. During this training phase, overfitting of the network was prevented by the use of L_2 -regularization technique (i.e. penalizing by adding the sum of the squared values of the weights to the error function) with parameter $\lambda = 0.1$ (Ng, 2004) and performance comparison against the validation set. The validation data was presented to the network after

each training iteration to obtain a non-training performance error. Training was stopped when the validation error ceased to decrease (Sarle, 1995). This training procedure was repeated for all networks constructed in the neuron grid search; performance of these networks was evaluated using the measures described below, and the optimal deep learning network size was selected.

2.8. Network evaluation

After the training stage, performance of the obtained neural network was assessed in the evaluation phase by means of the evaluation dataset. The input features (i.e. clinical characteristics (network 1), connectivity matrices (network 2), morphology values (network 3) or all three information sources (network 4)) of the subjects in the evaluation set (n = 32) were presented to the trained networks. The softmax activation function (Bishop, 1995) was used for the output nodes, resulting in a vector of values varying between 0 and 1 that add up to 1 and the output node with the highest probability was selected as the predicted class label using a winner-take-all approach (Duan et al., 2003; Hinton, 2002; Hinton et al., 2012; Lefebvre et al., 2013). To determine whether the predicted class was correct, the network output label was compared to the true class label (i.e. true survival class of a patient). Good classifications were marked by equal labels for the true and predicted class, incorrect classifications were marked by a mismatch between prediction and truth. Next, a mosaic plot of the trained network (Fig. 3) was computed to visualize the distribution of datasets over the different classes

after prediction (Hartigan and Kleiner, 1981), including the positive predictive values (PPV) of the network, defined as the percentage of patients with a predicted label that coincided with the true class label (Altman and Bland, 1994; Fletcher and Fletcher, 2005). For each class a PPV score was computed as $PPV_i = N_{correct}^i / N_{label\ i}$ ($i \in \{short, medium, long\}$), with $N_{correct}^i$ being the number of patients that correctly received class label i and $N_{label\ i}$ the total number of patients predicted to have class label i . PPVs were computed to give an impression of the discriminative power and thus predictive value of the network.

In addition, the overall performance of a network was assessed as the overall accuracy of the predictions, denoting the percentage of patients for whom the network predicted the correct class label and was calculated using the formula $N_{correct} / N$. Here, $N_{correct}$ denoted the number of patients for whom the prediction by the deep learning network was equal to the true survival class (i.e. highlighted diagonal elements in mosaic plot), and N denoted the total number of patients included in the set.

3. Results

3.1. Clinical deep learning

The clinical network was trained on eight clinical characteristics (site of onset, age at onset, time to diagnosis, ALSFRS slope, FVC, C9orf72 phenotype status, FTD status and El Escorial criteria diagnostic category). Neuron grid search resulted in a deep learning network with

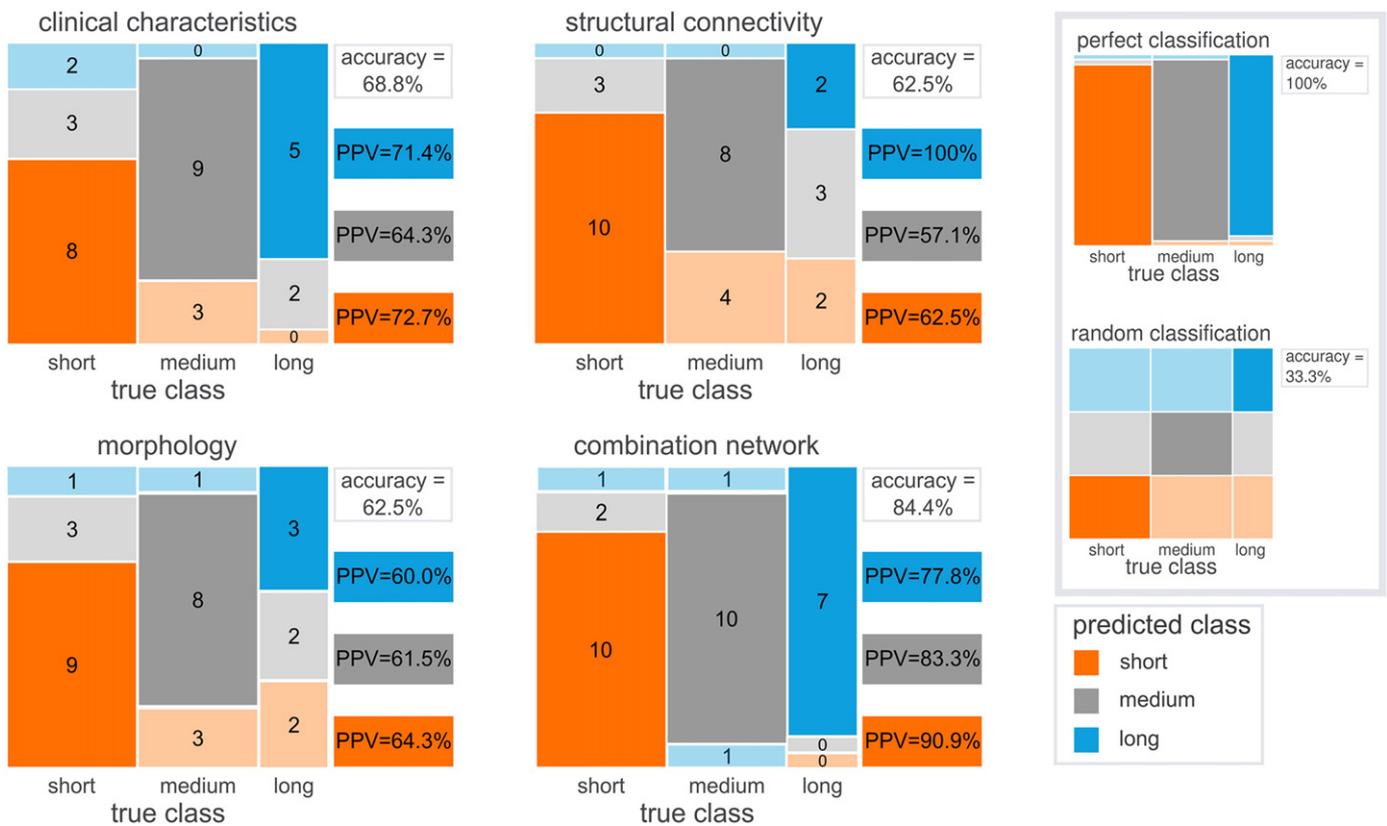


Fig. 3. The distribution of the prediction results shown in mosaic plots. The columns represent the known survival class of patients, where the width of columns is relative to the number of subjects in that column. The colors orange, gray and blue represent the predicted survival classes short, medium and long, respectively. The highlighted diagonal cells (bottom left to top right) denote the number of patients that were correctly classified and the off-diagonal cells the number of patients that were mispredicted (opaque cells), i.e. the wrong class was predicted by the network. The positive predictive value (Fletcher and Fletcher, 2005) for each predicted class can be derived by dividing the correctly predicted subjects by the total number of predictions of that class (i.e. highlighted cell/sum of same colored cells). The overall accuracy is computed by summing the highlighted diagonal cells and dividing this number by the total population. For the evaluation set (n = 32), the clinical deep learning network, the structural connectivity deep learning network, morphology deep learning network and the clinical-MRI combined deep learning network obtained overall accuracies of 68.8%, 62.5%, 62.5% and 78.1%, respectively. (Box) A perfect classification and a random classification mosaic plot are displayed for comparison of the results obtained on the evaluation set. In the perfect classification, all subjects are correctly classified and therefore an accuracy of 100% is achieved. In a random classification, three class labels are randomly distributed over the subjects, only predicting a third of the total subject population correctly.

8 input nodes, 158 nodes in the first hidden layer, 448 nodes in the second hidden layer, and 3 output nodes. The predicted short survivors had a PPV of 72.7% on the evaluation set; medium survivors had a PPV of 64.3% and long survivors 71.4%. The highest PPV was obtained for the predicted short survivors, indicating that a predicted short survival on the basis of clinical metrics was more often correct than a predicted medium or long survival. The optimal network based on clinical characteristics gave an evaluation accuracy of 68.8% (Fig. 3), a training accuracy of 78.3% (Fig. S1) and a validation accuracy of 70.0% (Fig. S2).

3.2. Structural connectivity deep learning

The structural connectivity MRI network, based on connection weights, consisted of 2285 input nodes (the total number of reconstructed tracts), and was fitted 134 and 313 nodes in the first and second hidden layer respectively, and comprised 3 output nodes. The PPV scores for the predicted short, medium and long survivors of the evaluation set were 62.5%, 57.1% and 100.0%, respectively. The highest PPV in this network was obtained for the long survivor class, indicating that this structural connectivity-based deep learning network was highly reliable when it gave predictions for the class of long survivors. The structural connectivity deep learning network reached an evaluation prediction accuracy of 62.5% (Fig. 3), a training accuracy of 79.5% (Fig. S1) and a validation accuracy of 60.0% (Fig. S2). With average simulated PPV chance levels of 38.6% (short), 38.6% (medium) and 23.0% (long) when assigning random class labels to patients, these findings show that objective connectivity values alone can provide valuable information on disease survival.

3.3. Morphology deep learning

The morphology MRI deep learning network consisted of 83 input nodes (68 cortical thickness values and 15 subcortical volumes), and was fitted to 181 and 178 nodes in the first and second hidden layer respectively, and contained included the three survival classes as output nodes. The PPV scores of the evaluations set were 64.3% (short), 61.5% (medium) and 60.0% (long survival), respectively. The highest PPV in this network was obtained for the short survivor class, indicating that

the morphology network, similar to the clinical network, was more reliable when it predicted a short survivor. The morphology deep learning network reached an evaluation prediction accuracy of 62.5% (Fig. 3), a training accuracy of 80.7% (Fig. S1) and a validation accuracy of 60.0% (Fig. S2). These findings support the relation between morphology and disease progression, and thus disease survival.

3.4. Clinical-MRI combined deep learning

Next, the prediction probabilities from the clinical deep learning network and the prediction probabilities from the two MRI networks were presented as input for a combined deep learning network. Grid search during network training resulted in a network configuration with 9 input nodes, 171 nodes in the first hidden layer, 108 nodes in the second hidden layer, and 3 output nodes. PPV scores of the combined network on the evaluation set included 90.9%, 83.3% and 77.8% for the predicted short, medium and long survivor classes, respectively. The combined network reached an evaluation accuracy of 84.4% (Fig. 3), a training accuracy of 88.0% (Fig. S1) and a validation accuracy of 80.0% (Fig. S2). Statistical testing indicated that survival prediction was significantly improved ($p < 0.001$) due to the addition of structural connectivity data and morphology MRI findings to the clinical characteristics (Fig. 4, see Supplementary materials for details).

4. Discussion

We evaluated the use of deep learning to predict survival time of ALS patients on the basis of clinical characteristics and advanced MRI metrics. Our findings show that MRI data alone (i.e. structural connectivity and brain morphology data, consisting of cortical thickness and subcortical volumes) can provide valuable predictions of survival time. Furthermore, combining clinical characteristics and MRI data into a layered deep learning approach can further improve predictions about whether a patient will have a short, medium or long survival time.

Previous studies have used proportional Cox hazard classification (Cox, 1972) employing clinical characteristics such as site of onset, executive dysfunction and diagnostic delay (Elamin et al., 2015; Scotton et al., 2012) to develop prognostic models for survival. These models already showed the predictive power of clinical data, with PPVs and overall accuracies lower than or similar to results in our study, indicating a potentially better predictive power of survival classes using deep learning. It should however be noted that PPV scores are dependent on the prevalence of a subtype in the total population and might also influence the differences in scores.

Deep learning on diffusion-weighted imaging data led to a prediction accuracy of 62.5%. Deep learning on T1-weighted images data resulted in a prediction accuracy of 62.5%. A combination of these imaging metrics yielded an improved prediction accuracy of 78.1% (see Supplementary materials for more details), indicating the predictive power of combining imaging metrics in deep learning. Previous studies have used MRI data for the prediction of diagnosis; that is, they used structural connectivity MRI data to differentiate between ALS patients and healthy controls, resulting in prediction accuracies between 70 and 80% (Fekete et al., 2013; Welsh et al., 2013). Other studies used cortical thickness measurements to discriminate between patients and controls in various diseases, such as Alzheimer's disease (Lerch et al., 2008), childhood onset schizophrenia (Greenstein et al., 2012), and major depression (Foland-Ross et al., 2015) with prediction accuracies ranging from 60 to 75%. In our study we examined a presumably more difficult task of predicting survival time within the group of patients, with a priori chance levels (i.e. true positive rate) here equal to 33.3% for the three survival classes, rather than 50% for patient/control status. The potential of MRI in patient classification and prognostication was previously also shown for the prediction of disease status in Alzheimer's disease, where machine learning differentiated between two subtypes of dementia based on T1-weighted images with accuracies of 89%

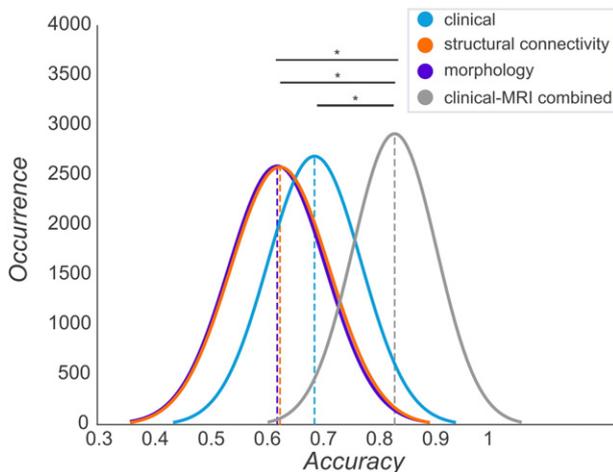


Fig. 4. Fitted normal curves on accuracy distributions of the four networks. Normal curves are fitted on the accuracy distributions of the clinical deep learning network (blue), the structural connectivity deep learning network (orange), morphology deep learning network (purple) and the clinical-MRI combined deep learning network (gray), based on 16 randomly selected subjects (repeated 10,000 times) from the first evaluation set ($n = 32$). The mean accuracies (dashed lines) of these distributions were 68.7%, 62.5%, 62.4% and 84.4% for the clinical deep learning network, structural connectivity deep learning network, morphology deep learning network and clinical-MRI combined deep learning network, respectively. A paired t -test showed significant differences between accuracies of each pair of networks (all $p < 0.001$).

(Klöppel et al., 2008). In this study we included all reconstructed connections instead of focusing on connections between specific brain regions, for example from motor regions to other brain areas (Schmidt et al., 2014). By considering all connections, a deep learning method is allowed to identify combinations of affected connections that are most valuable for survival prediction. As such, the deep learning network may detect relevant patterns in connections that are only slightly affected, thereby adding valuable information for prediction. The ability of deep learning to distill complex relationships from large datasets makes it a promising tool for disease prognostication.

The predictions of the network based on clinical parameters and the two MRI networks were combined in a clinical-MRI network. The clinical network seemed to be less sensitive to correctly predicting short survivors compared to the MRI networks (see Supplementary materials for more details). The combined network learned relationships between the survival class predictions of the three other networks. Patients incorrectly predicted by either the clinical or one of the MRI networks were often predicted correctly by the combination network using prediction information from the other two networks. By utilizing the predicting probabilities of the survival classes instead of the survival class label, the uncertainty of predictions was taken into account. The prediction probabilities of the clinical, structural connectivity MRI and morphology MRI networks contributed equally to the combined prediction.

Deep learning shows promising results for the prediction of survival categories for individual ALS patients, but several points have to be taken into account. Large training and evaluation sets are preferred to ensure convergence of prediction accuracies and to prevent overfitting (Ng, 2004). In addition, external validation is crucial for the development of a reliable prognostic tool and should be incorporated in future examinations. Second, while deep learning can effectively make predictions on datasets with complex relationships, dependency among input variables and between input and output variables cannot be easily deduced. In future research, it would therefore be worthwhile to investigate possibilities to reveal these dependencies and gain more insight into the mechanisms underlying disease progression. Third, prediction may also be improved using additional deep learning networks based on fMRI scans, as used in previous studies investigating disease diagnosis (Fekete et al., 2013; Welsh et al., 2013). Finally, additional clinical characteristics or diffusion tensor imaging metrics may also improve prediction. For example, radial diffusivity differences have been shown between ALS patients and healthy controls (Agosta et al., 2010; Metwalli et al., 2010) and therefore might also be of value in survival prognostication.

Deep learning is a powerful approach with successful applications in many real world issues. Here, we show that deep learning can also be of benefit to medical problems. Our findings show that deep learning can contribute to early prognostication of survival in ALS by combining clinical characteristics and brain imaging data. Our study provides promising results and may contribute to developing an automated prognostication tool for the estimation of survival in individual patients.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.nicl.2016.10.008>.

Disclosure

MPvdH was supported by the Netherlands Organization for Scientific Research VIDI Grant and a fellowship of the Brain Center Rudolf Magnus. LHvdB received funding from the Netherlands Organization for Scientific Research VICI Grant and from the ALS Foundation Netherlands. LHvdB received travel grants and consultancy fees from Baxter and serves on scientific advisory boards for Prinses Beatrix Spierfonds, Thierry Latran Foundation, Cytokinetics and Biogen Idec.

References

Agosta, F., Pagani, E., Petrolini, M., Caputo, D., Perini, M., Prella, A., Salvi, F., Filippi, M., 2010. Assessment of white matter tract damage in patients with amyotrophic lateral

- sclerosis: a diffusion tensor MR imaging tractography study. *AJNR Am. J. Neuroradiol.* 31:1457–1461. <http://dx.doi.org/10.3174/ajnr.A2105>.
- Agosta, F., Valsasina, P., Absinta, M., Riva, N., Sala, S., Prella, A., Copetti, M., Comola, M., Comi, G., Filippi, M., 2011. Sensorimotor functional connectivity changes in amyotrophic lateral sclerosis. *Cereb. Cortex* 21:2291–2298. <http://dx.doi.org/10.1093/cercor/bhr002>.
- Agosta, F., Valsasina, P., Riva, N., Copetti, M., Messina, M.J., Prella, A., Comi, G., Filippi, M., 2012. The cortical signature of amyotrophic lateral sclerosis. *PLoS One* 7, e42816. <http://dx.doi.org/10.1371/journal.pone.0042816>.
- Ahmed, B., Brodley, C.E., Blackmon, K.E., Kuzniecky, R., Barash, G., Carlson, C., Quinn, B.T., Doyle, W., French, J., Devinsky, O., Thesen, T., 2015. Cortical feature analysis and machine learning improves detection of "MRI-negative" focal cortical dysplasia. *Epilepsy Behav.* 48:21–28. <http://dx.doi.org/10.1016/j.yebeh.2015.04.055>.
- Alexander, A.L., Lee, J.E., Lazar, M., Field, A.S., 2007. Diffusion tensor imaging of the brain. *Neurotherapeutics* 4:316–329. <http://dx.doi.org/10.1016/j.nurt.2007.05.011>.
- Altman, D.G., Bland, J.M., 1994. Diagnostic tests 2: predictive values. *BMJ* 309:102. <http://dx.doi.org/10.1136/bmj.309.6947.102>.
- Andersson, J.L.R., Skare, S., 2002. A model-based method for retrospective correction of geometric distortions in diffusion-weighted EPI. *NeuroImage* 16:177–199. <http://dx.doi.org/10.1006/nimg.2001.1039>.
- Andersson, J.L.R., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage* 20:870–888. [http://dx.doi.org/10.1016/S1053-8119\(03\)00336-7](http://dx.doi.org/10.1016/S1053-8119(03)00336-7).
- Bishop, C.M., 1995. *Neural Networks for Pattern Recognition*. Oxford University Press, Inc. Bretschneider, J., Del Tredici, K., Toledo, J.B., Robinson, J.L., Irwin, D.J., Grossman, M., Suh, E., Van Deerlin, V.M., Wood, E.M., Baek, Y., Kwong, L., Lee, E.B., Elman, L., McCluskey, L., Fang, L., Feldengut, S., Ludolph, A.C., Lee, V.M., Braak, H., Trojanowski, J.Q., 2013. Stages of pTDP-43 pathology in amyotrophic lateral sclerosis. *Ann. Neurol.* 74: 20–38. <http://dx.doi.org/10.1002/ana.23937>.
- Brooks, B.R., Miller, R.G., Swash, M., Munsat, T.L., 2009. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler. Other Motor Neuron Disord.* 1:293–299. <http://dx.doi.org/10.1080/146608200300079536>.
- Bullmore, E.T., Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10:186–198. <http://dx.doi.org/10.1038/nrn2575>.
- Cedarbaum, J.M., Stambler, N., Malta, E., Fuller, C., Hilt, D., Thurmond, B., Nakanishi, A., BDNF ALS study group (Phase III), 1999. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. *J. Neurol. Sci.* 169: 13–21. [http://dx.doi.org/10.1016/S0022-510X\(99\)00210-5](http://dx.doi.org/10.1016/S0022-510X(99)00210-5).
- Chiò, A., Logrosino, G., Hardiman, O., Swingle, R., Mitchell, D., Beghi, E., Traynor, B.G., Eurals Consortium, 2009. Prognostic factors in ALS: a critical review. *Amyotroph. Lateral Scler.* 10:310–323. <http://dx.doi.org/10.3109/17482960802566824>.
- Chiò, A., Calvo, A., Moglia, C., Mazzini, L., Mora, G., PARALS study group, 2011. Phenotypic heterogeneity of amyotrophic lateral sclerosis: a population based study. *J. Neurol. Neurosurg. Psychiatry* 82:740–746. <http://dx.doi.org/10.1136/jnnp.2010.235952>.
- Ciccarelli, O., Behrens, T.E., Altmann, D.R., Orrell, R.W., Howard, R.S., Johansen-Berg, H., Miller, D.H., Matthews, P.M., Thompson, A.J., 2006. Probabilistic diffusion tractography: a potential tool to assess the rate of disease progression in amyotrophic lateral sclerosis. *Brain* 129:1859–1871. <http://dx.doi.org/10.1093/brain/awl100>.
- Cox, D.R., 1972. Regression models and life-tables. *J. R. Stat. Soc. Ser. B* 34:187–220. http://dx.doi.org/10.1007/978-1-4612-4380-9_37.
- Cristianini, N., Shawe-Taylor, J., 2000. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press, New York, New York, USA. <http://dx.doi.org/10.1017/CBO9780511801389>.
- Crowther, P.S., Cox, R.J., 2005. A method for optimal division of data sets for use in neural networks. In: Khosla, R., Howlett, R.J., Jain, L.C. (Eds.), *Knowledge-Based Intelligent Information and Engineering Systems: 9th International Conference, KES 2005, Melbourne, Australia, September 14–16, 2005, Proceedings, Part IV*. Springer, Berlin, Heidelberg:pp. 1–7. http://dx.doi.org/10.1007/11554028_1.
- Dean, J., Corrado, G.S., Monga, R., Chen, K., Devin, M., Le, Q.V., Mao, M.Z., Ranzato, M.A., Senior, A., Tucker, P., Yang, K., Ng, A.Y., 2012. *Large scale distributed deep networks*. *Adv. Neural Inf. Proces. Syst.* 25, 1232–1240.
- DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G.-Y.R., Karydas, A., Seeley, W.W., Josephs, K.A., Coppola, G., Geschwind, D.H., Wszolek, Z.K., Feldman, H., Knopman, D.S., Petersen, R.C., Miller, B.L., Dickson, D.W., Boylan, K.B., Graff-Radford, N.R., Rademakers, R., 2011. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTLD and ALS. *Neuron* 72:245–256. <http://dx.doi.org/10.1016/j.neuron.2011.09.011>.
- del Aguila, M.A., Longstreth, W.T., McGuire, V., Koepsell, T.D., van Belle, G., 2003. Prognosis in amyotrophic lateral sclerosis: a population-based study. *Neurology* 60:813–819. <http://dx.doi.org/10.1212/01.WNL.0000049472.47709.3B>.
- Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 31:968–980. <http://dx.doi.org/10.1016/j.neuroimage.2006.01.021>.
- Duan, K., Keerthi, S.S., Chu, W., Shevade, S.K., Poo, A.N., 2003. Multi-category classification by soft-max combination of binary classifiers. In: Windeatt, T., Roli, F. (Eds.), *Multiple Classifier Systems: 4th International Workshop, MCS 2003 Guildford, UK, June 11–13, 2003 Proceedings*. Springer, Berlin Heidelberg, Berlin, Heidelberg:pp. 125–134. http://dx.doi.org/10.1007/3-540-44938-8_13.
- Duda, R.O., Hart, P.E., Stork, D.G., 2001. *Pattern Classification*. John Wiley & Sons, New York, New York, USA.

- Elamin, M., Bede, P., Montuschi, A., Pender, N., Chio, A., Hardiman, O., 2015. Predicting prognosis in amyotrophic lateral sclerosis: a simple algorithm. *J. Neurol.* 262: 1447–1454. <http://dx.doi.org/10.1007/s00415-015-7731-6>.
- Fekete, T., Zach, N., Mujica-Parodi, L.R., Turner, M.R., 2013. Multiple kernel learning captures a systems-level functional connectivity biomarker signature in amyotrophic lateral sclerosis. *PLoS One* 8, e85190. <http://dx.doi.org/10.1371/journal.pone.0085190>.
- Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl. Acad. Sci. U. S. A.* 97:11050–11055. <http://dx.doi.org/10.1073/pnas.200033797>.
- Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33:341–355. [http://dx.doi.org/10.1016/S0896-6273\(02\)00569-X](http://dx.doi.org/10.1016/S0896-6273(02)00569-X).
- Fletcher, R.H., Fletcher, S.W., 2005. *Clinical Epidemiology: The Essentials*. 4th ed. Lippincott Williams & Wilkins.
- Foland-Ross, L.C., Sacchet, M.D., Prasad, G., Gilbert, B., Thompson, P.M., Gotlib, I.H., 2015. Cortical thickness predicts the first onset of major depression in adolescence. *Int. J. Dev. Neurosci.* 46:125–131. <http://dx.doi.org/10.1016/j.ijdevneu.2015.07.007>.
- Greenstein, D., Malley, J.D., Weisinger, B., Clasen, L., Gogtay, N., 2012. Using multivariate machine learning methods and structural MRI to classify childhood onset schizophrenia and healthy controls. *Front. Psych.* 3:53. <http://dx.doi.org/10.3389/fpsy.2012.00053>.
- Hardiman, O., van den Berg, L.H., Kiernan, M.C., 2011. Clinical diagnosis and management of amyotrophic lateral sclerosis. *Nat. Rev. Neurol.* 7:639–649. <http://dx.doi.org/10.1038/nrneuro.2011.153>.
- Hartigan, J.A., Kleiner, B., 1981. Mosaics for contingency tables. *Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface*. Springer:pp. 268–273. http://dx.doi.org/10.1007/978-1-4613-9464-8_37.
- Hinton, G.E., 2002. Training products of experts by minimizing contrastive divergence. *Neural Comput.* 14:1771–1800. <http://dx.doi.org/10.1162/089976602760128018>.
- Hinton, G.E., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., Kingsbury, B., 2012. Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Process. Mag.* 29:1–27. <http://dx.doi.org/10.1109/MSP.2012.2205597>.
- Hjelm, R.D., Calhoun, V.D., Salakhutdinov, R., Allen, E.A., Adali, T., Plis, S.M., 2014. Restricted Boltzmann machines for neuroimaging: an application in identifying intrinsic networks. *NeuroImage* 96:245–260. <http://dx.doi.org/10.1016/j.neuroimage.2014.03.048>.
- Karsoliya, S., 2012. Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture. *Int. J. Eng. Trends Technol.* 3, 714–717.
- Kassubek, J., Muller, H.P., Del Tredici, K., Brettschneider, J., Pinkhardt, E.H., Lule, D., Bohm, S., Braak, H., Ludolph, A.C., 2014. Diffusion tensor imaging analysis of sequential spreading of disease in amyotrophic lateral sclerosis confirms patterns of TDP-43 pathology. *Brain* 137:1733–1740. <http://dx.doi.org/10.1093/brain/awu090>.
- Kimura, F., Fujimura, C., Ishida, S., Nakajima, H., Furutama, D., Uehara, H., Shinoda, K., Sugino, M., Hanafusa, T., 2006. Progression rate of ALSFRS-R at time of diagnosis predicts survival time in ALS. *Neurology* 66:265–267. <http://dx.doi.org/10.1212/01.wnl.0000194316.91908.8a>.
- Klöppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scapellato, R.I., Rohrer, J.D., Fox, N.C., Jack, C.R., Ashburner, J., Frackowiak, R.S.J., 2008. Automatic classification of MR scans in Alzheimer's disease. *Brain* 131:681–689. <http://dx.doi.org/10.1093/brain/awn319>.
- Kollewe, K., Mauss, U., Krampfl, K., Petri, S., Dengler, R., Mohammadi, B., 2008. ALSFRS-R score and its ratio: a useful predictor for ALS-progression. *J. Neurol. Sci.* 275:69–73. <http://dx.doi.org/10.1016/j.jns.2008.07.016>.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. *ImageNet Classification With Deep Convolutional Neural Networks*. *Advances in Neural Information Processing Systems* 25 (NIPS 2012).
- Larochelle, H., Bengio, Y., Louradour, J., Lamblin, P., 2009. Exploring strategies for training deep neural networks. *J. Mach. Learn. Res.* 10, 1–40.
- Lefebvre, G., Berlemont, S., Mamalet, F., Garcia, C., 2013. BLSTM-RNN based 3D gesture classification. In: Mladenov, V., Koprinkova-Hristova, P., Palm, G., Villa, A.E.P., Appollini, B., Kasabov, N. (Eds.), *Artificial Neural Networks and Machine Learning (ICANN 2013)*. Springer, Berlin Heidelberg, Sofia, Bulgaria. <http://dx.doi.org/10.1007/978-3-642-40728-4>.
- Lerch, J.P., Pruessner, J., Zijdenbos, A.P., Collins, D.L., Teipel, S.J., Hampel, H., Evans, A.C., 2008. Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls. *Neurobiol. Aging* 29:23–30. <http://dx.doi.org/10.1016/j.neurobiolaging.2006.09.013>.
- Menke, R.A.L., Abraham, I., Thiel, C.S., Filipini, N., Knight, S., Talbot, K., Turner, M.R., 2012. Fractional anisotropy in the posterior limb of the internal capsule and prognosis in amyotrophic lateral sclerosis. *Arch. Neurol.* 69:1493–1499. <http://dx.doi.org/10.1001/archneurol.2012.1122>.
- Metwalli, N.S., Benatar, M., Nair, G., Usher, S., Hu, X., Carew, J.D., 2010. Utility of axial and radial diffusivity from diffusion tensor MRI as markers of neurodegeneration in amyotrophic lateral sclerosis. *Brain Res.* 1348:156–164. <http://dx.doi.org/10.1016/j.brainres.2010.05.067>.
- Mezzapesa, D.M., D'Errico, E., Tortelli, R., Distaso, E., Cortese, R., Tursi, M., Federico, F., Zoccollella, S., Logrosco, G., Dicuonzo, F., Simone, I.L., 2013. Cortical thinning and clinical heterogeneity in amyotrophic lateral sclerosis. *PLoS One* 8, e80748. <http://dx.doi.org/10.1371/journal.pone.0080748>.
- Mohamed, A.-r., Dahl, G.E., Hinton, G.E., 2012. Acoustic modeling using deep belief networks. *IEEE Trans. Audio Speech Lang. Process.* 20:14–22. <http://dx.doi.org/10.1109/TASL.2011.2109382>.
- Møller, M.F., 1993. A scaled conjugate gradient algorithm for fast supervised learning. *Neural Netw.* 6:525–533. [http://dx.doi.org/10.1016/S0893-6080\(05\)80056-5](http://dx.doi.org/10.1016/S0893-6080(05)80056-5).
- Mori, S., van Zijl, P.C.M., 2002. Fiber tracking: principles and strategies - a technical review. *NMR Biomed.* 15:468–480. <http://dx.doi.org/10.1002/nbm.781>.
- Mori, S., Crain, B.J., Chacko, V.P., van Zijl, P.C.M., 1999. Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Ann. Neurol.* 45: 265–269. [http://dx.doi.org/10.1002/1531-8249\(199902\)45:2<265::AID-ANA21>3.0.CO;2-3](http://dx.doi.org/10.1002/1531-8249(199902)45:2<265::AID-ANA21>3.0.CO;2-3).
- Mori, S., Kaufmann, W.E., Davatzikos, C., Stieltjes, B., Amodei, L., Fredericksen, K., Pearlson, G.D., Melhem, E.R., Solaiyappan, M., Raymond, G.V., Moser, H.W., van Zijl, P.C.M., 2002. Imaging cortical association tracts in the human brain using diffusion-tensor-based axonal tracking. *Magn. Reson. Med.* 47:215–223. <http://dx.doi.org/10.1002/mrm.10074>.
- Müller, H.P., Turner, M.R., Grosskreutz, J., Abrahams, S., Bede, P., Govind, V., Prudlo, J., Ludolph, A.C., Filippi, M., Kassubek, J., Neuroimaging Society in, A.L.S.D.T.I.S.G., 2016. A large-scale multicentre cerebral diffusion tensor imaging study in amyotrophic lateral sclerosis. *J. Neurol. Neurosurg. Psychiatry* 87:570–579. <http://dx.doi.org/10.1136/jnnp-2015-311952>.
- Murphy, K.P., 2012. *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, Mass.
- Neary, D., Snowden, J.S., Gustafson, L., Passant, U., Stuss, D., Black, S., Freedman, M., Kertesz, A., Robert, P.H., Albert, M., Boone, K., Miller, B.L., Cummings, J., Benson, D.F., 1998. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology* 51:1546–1554. <http://dx.doi.org/10.1212/WNL.51.6.1546>.
- Ng, A.Y., 2004. Feature selection, L1 vs. L2 regularization, and rotational invariance. *ICML '04 Proceedings of the Twenty-First International Conference on Machine Learning*. ACM Press: p. 78. <http://dx.doi.org/10.1145/1015330.1015435>.
- Plis, S.M., Hjelm, D.R., Salakhutdinov, R., Allen, E.A., Bockholt, H.J., Long, J.D., Johnson, H.J., Paulsen, J.S., Turner, J.A., Calhoun, V.D., 2014. Deep learning for neuroimaging: a validation study. *Front. Neurosci.* 8:229. <http://dx.doi.org/10.3389/fnins.2014.00229>.
- Priddy, K.L., Keller, P.E., 2005. *Data Normalization*. *Artificial Neural Networks: An Introduction*. SPIE - The International Society for Optical Engineering, Bellingham, Washington, pp. 15–17.
- Qureshi, M.M., Hayden, D., Urbinelli, L., Ferrante, K., Newhall, K., Myers, D., Hilgenberg, S., Smart, R., Brown, R.H., Cudkowicz, M.E., 2006. Analysis of factors that modify susceptibility and rate of progression in amyotrophic lateral sclerosis (ALS). *Amyotroph. Lateral Scler.* 7:173–182. <http://dx.doi.org/10.1080/14660820600640596>.
- Rascovsky, K., Hodges, J.R., Knopman, D., Mendez, M.F., Kramer, J.H., Neuhaus, J., van Swieten, J.C., Seelaar, H., Dopper, E.G., Onyike, C.U., Hillis, A.E., Josephs, K.A., Boeve, B.F., Kertesz, A., Seeley, W.W., Rankin, K.P., Johnson, J.K., Gorno-Tempini, M.L., Rosen, H., Priloleau-Latham, C.E., Lee, A., Kipps, C.M., Lillo, P., Piguet, O., Rohrer, J.D., Rossor, M.N., Warren, J.D., Fox, N.C., Galasko, D., Salmon, D.P., Black, S.E., Mesulam, M., Weintraub, S., Dickerson, B.C., Diehl-Schmid, J., Pasquier, F., Deramecourt, V., Lebert, F., Pijnenburg, Y., Chow, T.W., Manes, F., Grafman, J., Cappa, S.F., Freedman, M., Grossman, M., Miller, B.L., 2011. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain* 134:2456–2477. <http://dx.doi.org/10.1093/brain/awr179>.
- Ravits, J.M., La Spada, A.R., 2009. ALS motor phenotype heterogeneity, focal, and spread: deconstructing motor neuron degeneration. *Neurology* 73:805–811. <http://dx.doi.org/10.1212/WNL.0b013e3181b6bbbd>.
- Renton, A.E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas, L., Kalimo, H., Paetau, A., Abramzon, Y., Remes, A.M., Kaganovich, A., Scholz, S.W., Duckworth, J., Ding, J., Harmer, D.W., Hernandez, D.G., Johnson, J.O., Mok, K., Ryten, M., Trabzuni, D., Guerreiro, R.J., Orrell, R.W., Neal, J., Murray, A., Pearson, J., Jansen, I.E., Sondervan, D., Seelaar, H., Blake, D., Young, K., Halliwell, N., Callister, J.B., Toulson, G., Richardson, A., Gerhard, A., Snowden, J., Mann, D., Neary, D., Nalls, M.A., Peuralinna, T., Jansson, L., Isoviita, V.M., Kaivorinne, A.L., Holtta-Vuori, M., Ikonen, E., Sulkava, R., Benatar, M., Wu, J., Chio, A., Restagno, G., Borghero, G., Sabatelli, M., Consortium, I., Heckerman, D., Rogava, E., Zinman, L., Rothstein, J.D., Sendtner, M., Drepper, C., Eichler, E.E., Alkan, C., Abdullaev, Z., Pack, S.D., Dutra, A., Pak, E., Hardy, J., Singleton, A., Williams, N.M., Heutink, P., Pickering-Brown, S., Morris, H.R., Tienari, P.J., Traynor, B.J., 2011. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 72:257–268. <http://dx.doi.org/10.1016/j.neuron.2011.09.010>.
- Rubinstein, R.Y., Kroese, D.P., 2004. *The cross-entropy method: a unified approach to combinatorial optimization*. Monte-Carlo Simulation and Machine Learning. Springer, New York. <http://dx.doi.org/10.1007/978-1-4757-4321-0>.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. *Nature* 323:533–536. <http://dx.doi.org/10.1038/323533a0>.
- Sarle, W.S., 1995. *Stopped training and other remedies for overfitting*. *Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics*, pp. 352–360.
- Schmidt, R., Verstraete, E., de Reus, M.A., Veldink, J.H., van den Berg, L.H., van den Heuvel, M.P., 2014. Correlation between structural and functional connectivity impairment in amyotrophic lateral sclerosis. *Hum. Brain Mapp.* 35:4386–4395. <http://dx.doi.org/10.1002/hbm.22481>.
- Schmidt, R., de Reus, M.A., Scholtens, L.H., van den Berg, L.H., van den Heuvel, M.P., 2015. Simulating disease propagation across white matter connectome reveals anatomical substrate for neuropathology staging in amyotrophic lateral sclerosis. *NeuroImage* 124:762–769. <http://dx.doi.org/10.1016/j.neuroimage.2015.04.005>.
- Scotton, W.J., Scott, K.M., Moore, D.H., Almedom, L., Wijesekera, L.C., Janssen, A., Nigro, C., Sakel, M., Leigh, P.N., Shaw, C., Al-Chalabi, A., 2012. Prognostic categories for amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler.* 13:502–508. <http://dx.doi.org/10.3109/17482968.2012.679281>.
- Senda, J., Kato, S., Kaga, T., Ito, M., Atsuta, N., Nakamura, T., Watanabe, H., Tanaka, F., Naganawa, S., Sobue, G., 2011. Progressive and widespread brain damage in ALS: MRI voxel-based morphometry and diffusion tensor imaging study. *Amyotroph. Lateral Scler.* 12:59–69. <http://dx.doi.org/10.3109/17482968.2010.517850>.

- Shahin, M.A., Maier, H.R.J., Mark, B., 2004. Data division for developing neural networks applied to geotechnical engineering. *J. Comput. Civ. Eng.* 18:105–114. [http://dx.doi.org/10.1061/\(ASCE\)0887-3801\(2004\)18:2\(105\)](http://dx.doi.org/10.1061/(ASCE)0887-3801(2004)18:2(105)).
- Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., Hassabis, D., 2016. Mastering the game of Go with deep neural networks and tree search. *Nature* 529:484–489. <http://dx.doi.org/10.1038/nature16961>.
- Simon, N.G., Turner, M.R., Vucic, S., Al-Chalabi, A., Shefner, J., Lomen-Hoerth, C., Kiernan, M.C., 2014. Quantifying disease progression in amyotrophic lateral sclerosis. *Ann. Neurol.* 76:643–657. <http://dx.doi.org/10.1002/ana.24273>.
- Turner, M.R., Verstraete, E., 2015. What does imaging reveal about the pathology of amyotrophic lateral sclerosis? *Curr. Neurol. Neurosci. Rep.* 15:45. <http://dx.doi.org/10.1007/s11910-015-0569-6>.
- Turner, M.R., Grosskreutz, J., Kassubek, J., Abrahams, S., Agosta, F., Benatar, M., Filippi, M., Goldstein, L.H., van den Heuvel, M.P., Kalra, S., Lulé, D., Mohammadi, B., 2011. Towards a neuroimaging biomarker for amyotrophic lateral sclerosis. *Lancet Neurol.* 10:400–403. [http://dx.doi.org/10.1016/s1474-4422\(11\)70049-7](http://dx.doi.org/10.1016/s1474-4422(11)70049-7).
- van den Heuvel, M.P., Sporns, O., 2011. Rich-club organization of the human connectome. *J. Neurosci.* 31:15775–15786. <http://dx.doi.org/10.1523/JNEUROSCI.3539-11.2011>.
- van den Heuvel, M.P., Mandl, R.C., Luigjes, J., Hulshoff Pol, H.E., 2008. Microstructural organization of the cingulum tract and the level of default mode functional connectivity. *J. Neurosci.* 28:10844–10851. <http://dx.doi.org/10.1523/JNEUROSCI.2964-08.2008>.
- van den Heuvel, M.P., Kahn, R.S., Goñi, J., Sporns, O., 2012. High-cost, high-capacity backbone for global brain communication. *Proc. Natl. Acad. Sci. U. S. A.* 109:11372–11377. <http://dx.doi.org/10.1073/pnas.1203593109>.
- Verstraete, E., van den Heuvel, M.P., Veldink, J.H., Blanken, N., Mandl, R.C., Hulshoff Pol, H.E., van den Berg, L.H., 2010. Motor network degeneration in amyotrophic lateral sclerosis: a structural and functional connectivity study. *PLoS One* 5, e13664. <http://dx.doi.org/10.1371/journal.pone.0013664>.
- Verstraete, E., Veldink, J.H., Mandl, R.C.W., van den Berg, L.H., van den Heuvel, M.P., 2011. Impaired structural motor connectome in amyotrophic lateral sclerosis. *PLoS One* 6, e24239. <http://dx.doi.org/10.1371/journal.pone.0024239>.
- Verstraete, E., Veldink, J.H., Hendrikse, J., Schelhaas, H.J., van den Heuvel, M.P., van den Berg, L.H., 2012. Structural MRI reveals cortical thinning in amyotrophic lateral sclerosis. *J. Neurol. Neurosurg. Psychiatry* 83:383–388. <http://dx.doi.org/10.1136/jnnp-2011-300909>.
- Verstraete, E., Veldink, J.H., van den Berg, L.H., van den Heuvel, M.P., 2014. Structural brain network imaging shows expanding disconnection of the motor system in amyotrophic lateral sclerosis. *Hum. Brain Mapp.* 35:1351–1361. <http://dx.doi.org/10.1002/hbm.22258>.
- Walhout, R., Westeneng, H.-J., Verstraete, E., Hendrikse, J., Veldink, J.H., van den Heuvel, M.P., van den Berg, L.H., 2015. Cortical thickness in ALS: towards a marker for upper motor neuron involvement. *J. Neurol. Neurosurg. Psychiatry* 86:288–294. <http://dx.doi.org/10.1136/jnnp-2013-306839>.
- Welsh, R.C., Jelsone-Swain, L.M., Foerster, B.R., 2013. The utility of independent component analysis and machine learning in the identification of the amyotrophic lateral sclerosis diseased brain. *Front. Hum. Neurosci.* 7:251. <http://dx.doi.org/10.3389/fnhum.2013.00251>.
- Westeneng, H.-J., Verstraete, E., Walhout, R., Schmidt, R., Hendrikse, J., Veldink, J.H., van den Heuvel, M.P., van den Berg, L.H., 2015. Subcortical structures in amyotrophic lateral sclerosis. *Neurobiol. Aging* 36:1075–1082. <http://dx.doi.org/10.1016/j.neurobiolaging.2014.09.002>.
- Wolf, J., Safer, A., Wöhrle, J.C., Palm, F., Nix, W.A., Maschke, M., Grau, A.J., 2014. Factors predicting one-year mortality in amyotrophic lateral sclerosis patients – data from a population-based registry. *BMC Neurol.* 14:197. <http://dx.doi.org/10.1186/s12883-014-0197-9>.
- Wolf, J., Safer, A., Wöhrle, J.C., Palm, F., Nix, W.A., Maschke, M., Grau, A.J., 2015. Factors predicting survival in ALS patients – data from a population-based registry in Rhineland-Palatinate, Germany. *Neuroepidemiology* 44:149–155. <http://dx.doi.org/10.1159/000381625>.
- Wu, J., Yu, Y., Huang, C., Yu, K., 2015. Deep multiple instance learning for image classification and auto-annotation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA. <http://dx.doi.org/10.1109/CVPR.2015.7298968>.