Analysis of convolutional neural networks for fronto-temporal dementia biomarker discovery

Alfonso Estudillo-Romero, Raffaella Migliaccio, Bénédicte Batrancourt, Pierre Jannin, John S H Baxter

Analysis of Convolutional Neural Networks for Fronto-Temporal Dementia Biomarker Discovery

Alfonso Estudillo Romero[1], Raffaella Migliaccio[2], Bénédicte Batrancourt[2], Pierre Jannin[1], John S. H. Baxter[1*]

[1*]Laboratoire Traitement du Signal et de l'Image (LTSI, INSERM UMR 1099), Université de Rennes, Rennes, France.
[2]Frontal Functions and Pathology Laboratory (FrontLab), Institut du Cerveau, Paris, France.

*Corresponding author(s). E-mail(s): john.baxter@univ-rennes.fr;
Contributing authors: alfonso.estudillo@univ-rennes.fr;
lara.migliaccio@gmail.com; benedicte.batrancourt@upmc.fr;
pierre.jannin@univ-rennes.fr;

Abstract

Purpose: Frontotemporal lobe dementia (FTD) results from the degeneration of the frontal and temporal lobes. It can manifest in several different ways, leading to the definition of variants characterised by their distinctive symptomatologies. As these variants are detected based on their symptoms, it can be unclear if they represent different types of FTD or different symptomatological axes. The goal of this paper is to investigate this question with a constrained cohort of FTD patients in order to see if the heterogeneity within this cohort can be inferred from medical images rather than symptom severity measurements.
Methods: An ensemble of convolutional neural networks (CNNs) are used to classify diffusion tensor images collected from two databases consisting of 72 patients with behavioural variant FTD and 120 healthy controls. FTD biomarkers were found using voxel-based analysis on the sensitivities of these CNNs. Sparse principal components analysis (sPCA) is then applied on the sensitivities arising from the patient cohort in order to identify the axes along which the patients express these biomarkers. Finally, this is correlated with their symptom severity measurements in order to interpret the clinical presentation of each axis.
Results: The CNNs result in sensitivities and specificities between 83% and 92%. As expected, our analysis determines that all the robust biomarkers arise from the frontal and temporal lobes. sPCA identified four axes in terms of biomarker expression which are correlated with symptom severity measurements.

Conclusion: Our analysis confirms that behavioural variant FTD is not a singular type or spectrum of FTD, but rather that it has multiple symptomatological axes that relate to distinct regions of the frontal and temporal lobes. This analysis suggests that medical images can be used to understand the heterogeneity of FTD patients and the underlying anatomical changes that lead to their different clinical presentations.

Keywords: Frontotemporal lobe dementia, convolutional neural networks, voxel-based diktiometry, disease sub-typing

1 Introduction

Frontotemporal lobe dementia (FTD) is a collection of various neurodegenerative patterns that entail the progressive degradation of both cognitive functions (such as memory) and behavioural functioning due to the degradation of the frontal and temporal lobes [1]. FTD is a leading form of dementia-like disorder along with Alzhiemer's disease, vascular dementia, and Lewy body dementia, all of which are growing in prevalence with an ageing population. FTD is very heterogeneous with clinical variants including the behavioural variant (bvFTD), non-fluent variant primary progressive aphasia, and semantic-variant primary progressive aphasia [1].

However, sub-typing of FTD to this degree requires the variant's characteristic symptoms be noticeable and measurable, making it difficult to discern prior to their manifestation. Medical imaging, notably MRI, has been investigated as a potential source for biomarkers that could precede FTD's more overt symptoms and potential aid in distinguishing between its various sub-types. In addition, the symptoms that define these variants are not mutually exclusive and could lead to a patient being diagnosed with a particular variant when they actually display multiple ones. This is consistent with genetic evidence that shows a degree of overlap between variants [2].

Structural MRI has long been used to quantify cortical thinning as a marker of cortical degeneration which is the most characteristic sign of FTD [3]. These methods include those based on explicitly segmenting specific cortical regions in MRI [4] as well as voxel-based morphometry [5]. DTI-based methods use features such as the fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AxD) and radial diffusivity (RD) computed from tract-based spatial statistics in combination with functional connectivity features obtained from fMRI data have been able to classify Alzheimer's disease from bvFTD [6]. A similar method using an elastic net regression model attempted to distinguish presymptomatic FTD mutation carriers from controls suggested that the FTD-related pathological processes start in the white matter [7].

Deep learning shows a large capacity to detect diseases based on medical images, although research attention in dementia has been predominantly directed towards Alzheimer's disease [8]. These machine learning methods have been highly varied with some using segmentation to measure particular features (such as ventricular volume) as input to a neural network whereas others directly process the image as a whole without pre-specified features [9]. Unlike the aforementioned methods, few machine learning

tools permit explainable biomarkers that could help improve our understanding of FTD and offer features for further investigation and potential FTD patient staging. However, they frequently report accuracies, sensitivities, and specificities on the order of 80% which is a much higher threshold than simple statistical significance used in traditional neuroimage processing methods for the same classification problems [8]. This illustrates a fundamental gap between these two paradigms: the former may not immediately provide the accuracy required to be individually diagnostic but add to our understanding of FTD and the latter is the opposite.

Voxel-based diktiometry (VBD) is a recently-proposed method to bridge this divide for neurological disorders [10, 11]. This approach combines convolutional neural network (CNN) ensembles with voxel-based analysis of their sensitivity maps to detect robust and repeatable features for a particular classification or regression task. The sensitivity maps are calculated from the set of testing images used in repeated cross-validation of the CNN architecture. This allows for a larger number of independent CNNs to be trained. Voxel-based analysis (VBA) is then used to statistically isolate patterns that robustly appear across patients and random network initialisations.

Unlike traditional VBA, these patterns do not necessarily reflect a local region, but rather how it contributes to the understanding of the entire image used by the CNN. Thus, these biomarkers highlight not only correlations between the underlying variables at a specific region and the disorder of interest, but also those between patterns connecting multiple regions. One example would be asymmetry biomarkers which can be more indicative than unilateral ones [10]. These are fundamentally non-local, involving multiple regions (i.e. the same anatomy but on different hemispheres) unlike the local biomarkers found in traditional VBA. One limitation of this non-locality is that it is not always clear if two regions of the brain identified through VBD are related to a single complex biomarker or multiple simpler biomarkers which is particularly when there is significant patient variability.

Contributions

This article applies VBD specifically to bvFTD. To overcome the aforementioned limitation, the VBD results will be decomposed using sparse principal components analysis (sPCA). This allows us to extract biomarkers which are expressed differently in different patients and correlate with particular clinical scores. This framework is not meant to be a diagnostic tool *per se*, but a knowledge discovery tool for more targeted biomarkers sensitive to disease heterogeneity.

2 Materials and methods

2.1 Participants

We have used two datasets compiled from healthy individuals and FTD patients diagnosed with bvFTD. A summary is given in Table 1.

The first includes 21 FTD patients diagnosed with bvFTD and 18 healthy controls (HC) recruited for the ECOCAPTURE protocol ([12, 13], Clinicaltrials.gov:NTC02496312, NCT03272230) at the ICM, Salpêtrière hospital Paris. Included

patients must meet the following criterion: a) diagnosed with bvFTD according to Rascovsky's international criteria. The inclusion criteria should also meet b) no evidence of any other cerebral pathology; c) a MMSE superior or equal than 10; d) aged between 40 and 85; e) without evidence of any psychiatric condition and a MADRS score inferior to 20; f) no evidence of excessive consumption of psychotropic drugs; g) no major physical disability disrupting mobility; h) no heart pacemaker. The HC

Table 1: Participant demographic and clinical data comparing FTD patients and HC.

ECOCAPTURE	FTD patients ($n = 21$)	HC ($n = 18$)	Statistics	
Women	($n = 7$)	($n = 10$)	$\chi^2 = 0.26$	$p = 0.608$
Age	65.12 (9.23)	62.61 (7.24)	-0.897	$p = 0.376$ [a]
Anxiety and depression				
HADS	13.48 (7.75)	5.44 (2.87)	-3.719	
MADRS	11.90 (5.96)	2.11 (2.40)	-4.578	
Apathy				
DAS	31.57 (9.23)	19.78 (8.14)	-4.198	$p < 0.0002$ [a]
SAS	16.19 (4.88)	5.78 (3.12)	-7.780	$p < 0.0001$ [a]
Behavioural metrics				
Compulsivity	($n = 14$) 9.07 (15.51)	($n = 17$) 0.24 (0.97)	-3.409	$p = 0.0007$ [b]
Social disinhibition	($n = 14$) 8.64 (6.87)	($n = 17$) 3.47 (5.06)	-2.521	$p = 0.0117$ [b]
Cognition				
FAB	12.00 (3.48)	17.33 (0.84)	5.141	$p < 0.0001$ [b]
HAY B-A	($n = 20$) 96.48 (113.31)	($n = 18$) 36.97 (17.91)	-2.317	$p = 0.0312$ [c]
HAY error score	($n = 20$) 19.10 (12.60)	($n = 18$) 3.11 (2.56)	-4.648	$p < 0.0001$ [b]
MATTIS DRS	118.48 (9.60)	142.17 (1.29)	5.324	$p < 0.0001$ [b]
MMSE	23.71 (2.61)	29.39 (0.78)	5.127	$p < 0.0001$ [b]
Social and emotions				
Mini-SEA faux pas	($n = 18$) 9.27 (3.69)	($n = 18$) 13.46 (1.10)	4.615	$p < 0.0002$ [c]
Mini-SEA recognition	9.18 (2.27)	12.95 (0.95)	6.936	$p < 0.0001$ [c]
FTLDNI	FTD patients ($n = 51$)	HC ($n = 112$)	Statistics	
Women	($n = 16$)	($n = 65$)	$\chi^2 = 8.93$	$p = 0.00281$
Age	61.06 (6.61)	63.28 (7.81)	1.761	$p = 0.080160$ [a]
Cognition				
CDR	($n = 49$) 1.21 (0.65)	($n = 111$) 0.01 (0.08)	-9.995	$p < 0.0001$ [b]
MMSE	($n = 47$) 23.15 (4.80)	($n = 106$) 29.40 (0.08)	9.181	$p < 0.0001$ [b]
MOCA	($n = 33$) 16.85 (7.30)	($n = 60$) 27.92 (1.61)	7.227	$p < 0.0001$ [b]
Language				
Calif. Verb. Learn.	($n = 47$) 19.02 (8.18)	($n = 59$) 30.07 (4.07)	6.903	$p < 0.0001$ [b]
Neuropsychiatric Invent.				
Motor Disturbance	($n = 32$) 2.19 (0.59)	n/a		

n = count of available data when different than the total, mean (standard deviation), χ^2 = chi-squared. Abbreviations: DAS, Dimensional Apathy Scale; FAB, Frontal Assessment Battery; HADS, Hospital Anxiety and Depression Scale; HAY, Hayling test; MADRS, Montgomery-Asberg Depression Rating Scale; MATTIS, Mattis Dementia Rating Scale; Mini-SEA: minisocial & emotional assessment; MMSE, Mini Mental State Examination; SAS, Starkstein Apathy Scale; CDR, Clinical Dementia Rating Scale; MOCA, Montreal Cognitive Assessment - Total score with adjustment for education level; Calif. Verb. Learn, California Verbal Learning Test - Total correctly recalled items over four learning trials.

[a] Student t-test

[b] Wilcoxon test

[c] Welch test

group must meet criteria d) - h) for matching purposes. All participants underwent the same cognitive, behavioural, and apathy tests. The MRI acquisition protocol used a 3T Siemens MRI scanner 64-channel Tim system and included T1 MPRAGE scans (TR = 2400 ms, TE = 2.17 ms, matrix = 320×320×256, slice thickness = 0.7 mm) and single-shot spin-echo diffusion weighted images (DWI) with 60 directions, covering the whole head with a posterior-anterior phase acquisition (b0 = 0 s/mm^2, b = 2000 s/mm^2, TE = 75 ms, TR = 3500 ms, flip angle = 90 degrees, field of view = 208 mm^2, voxel size = 1.76 × 1.76 × 1.76 mm^3). All acquisitions were performed at CENIR (Human MRI Neuroimaging core facility, Salpêtrière hospital, Paris, France)

The second dataset was collected at three different sites from the Fronto-Temporal Lobar Degeneration Neuroimaging Initiative (FTLDNI) [14]. The dataset comprises 121 HC as well as patients with different FTD variants of which we use the 51 with bvFTD. Recruited patients had neurological history and examination, collateral source interview, and neuropsychological testing. The clinical measures mostly come from the third version of the NIH National Alzheimer's Coordinating Center's (NACC) Uniform Data Set neuropsychological battery, which includes a module for assessment of FTD [14]. Behaviour, comportment, personality, language aspects and Clinical Dementia Rating Scale (CDR) scoring was used for staging [14]. We included all patients diagnosed with bvFTD and HC. The inclusion criteria for both groups was to have both T1 and DWI sequences. The MRI acquisitions for the selected individuals were performed at the Neuroscience Imaging Center at the University of California, San Francisco and at the Athinoula A. Martinos Center for Biomedical Imaging at Massachusetts General Hospital. The MRI acquisition protocol used a 3T Siemens MRI scanner 12-channel Tim Trio system. T1 MPRAGE scans used TR = 2300 ms, TE = 2.9 ms, matrix = 240×256×160 and a slice thickness = 1 mm. DWI data was acquired using a single-shot spin-echo sequence with 64 directions, covering the whole head with a posterior-anterior phase acquisition (b0 = 0 s/mm^2, b = 2000 s/mm^2, TE = 86 ms, voxel size = 2.2 × 2.2 × 2.2 mm^3). The TR was either 8200 or 6600 ms and the flip angle was 90 or 180 degrees depending on the acquisition centre.

2.2 Population-wise registration

Standard preprocessing tools were used for noise removal and bias field correction on the T1 images [15, 16] and for noise removal, ringing artifacts removal and bias field correction on the DWI images [17]. Also for the DWI sequences, the Eddy currents and subject motion artifacts were corrected using FSL [18]. We then calculated the DTI by means of weighted least squares (WLS) fitting method included in SlicerDMRI [19].

Two co-registrations were calculated for each subject using the Advanced Neuroimaging Tools [20, 21] (ANTs). The first one, a rigid registration between the non diffusion weighted image b0 and the T1 image and the second one a deformable registration between the T1 image and the MNI template [22]. This registration permits the standardisation of the images before being provided to the CNNs and ensures all sensitivity information is given in the common MNI co-ordinate space.

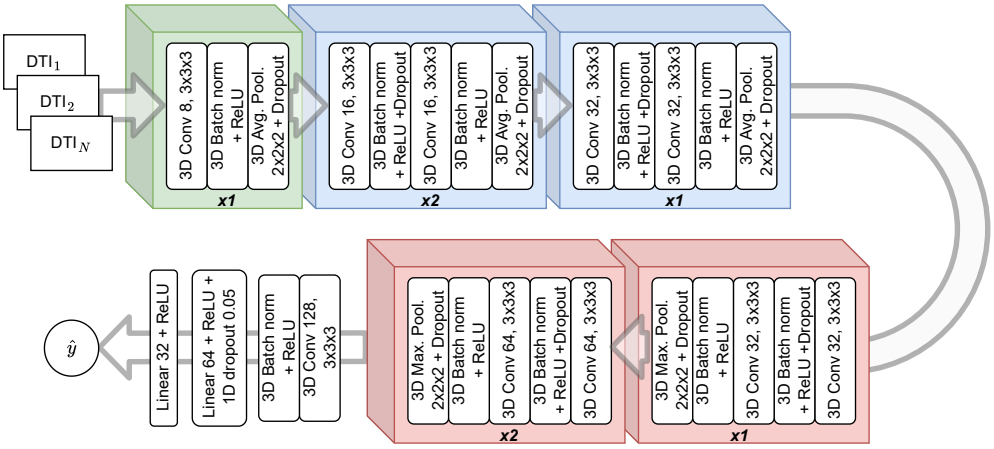


Fig. 1: Convolutional neural network architecture.

2.3 CNN training and testing

We have trained a simple CNN architecture as shown in Figure 1 Given the small amount of data, 4-fold cross-validation was used to estimate the CNN's performance and create the CNN ensemble. We trained each CNN for 120 epochs with a starting learning rate (l_r) of 2×10^{-4} that gradually decreased 50% every 30 epochs to finally end the training with a $l_r = 2.5 \times 10^{-5}$. We used Adam optimiser and a L2 regularisation of 0.5 for the CNN optimisation and binary cross-entropy as the loss function. Whether or not the networks contain information about the relevant classification problem is determined via a χ^2-test. Specifically, if the p-value is less than 5×10^{-4}, then we considered the network to be valid for biomarker discovery.

2.4 Voxel-based analysis statistical tests and filtering

We calculated the sensitivity maps of the ensemble of CNNs relative to the mean diffusivity (MD), anisotropy (A), fractional anisotropy (FA) and pseudo-planarity (PsPl). Since all sensitivity maps were already registered in the MNI space, we directly performed a two sample t-test where the first and second groups consisted of the DFT and HC cohorts, respectively. Finally a significance threshold of $p < 0.05$ FWE was applied to obtain the positive and negative correlated diktiometry maps within a minimum cluster size of 256 voxels. Apart from the analysis through VBD we also performed a VBA on the traditional diffusivity measures.

2.5 Sparse PCA

One of the limitations of VBD for disease classification is that it couples together potentially multiple distinct biomarkers [10] which is of particular importance for highly heterogeneous diseases such as FTD. In order to decompose the result into potentially distinct, we decomposed the sensitivity maps using sparse principal components analysis (sPCA) [23]. The aim of sPCA is to represent a set of features in a

Table 2: Confusion matrix, sensitivity, specificity, and balanced accuracy of the CNNs on the testing folds.

		Ground truth	
		HC	FTD
Prediction	HC	120	12
	FTD	12	60
Sensitivity:			83.33 %
Specificity:			92.30 %
BACC:			87.8%
χ^2**-statistic:**			113.76
p**-value:**			$< 1 \times 10^{-5}$

lower-dimensional space by computing its n principal components through a sparse dictionary that penalises those with low variance. This decomposed the sensitivity map into a series of maps that are not necessary orthogonal but express when signals in these regions frequently co-occur as a group. These are often spatial clusters although this is not guaranteed. We thus retain only spatial clusters with a minimum size of 256 mm^3 across the sensitivities to remove these smaller, likely spurious results. We set the parameter controlling the sparsity α equal to 0.8 and calculated $n = 5$ components.

sPCA was applied to the patient-specific sensitivity maps MD, A and PsPl specifically on the FTD patients since FA depends on MD and A [10]. Linear regression was used to see if the particular principal components related to particular symptomatology measures which could facilitate their interpretation. It is notable that this symptomatology information was not used for developing VBD biomarkers, only the knowledge of whether or not the patient has bvFTD. Thus, any correlation with clinical scores would show biomarkers for bvFTD that explain that particular symptom.

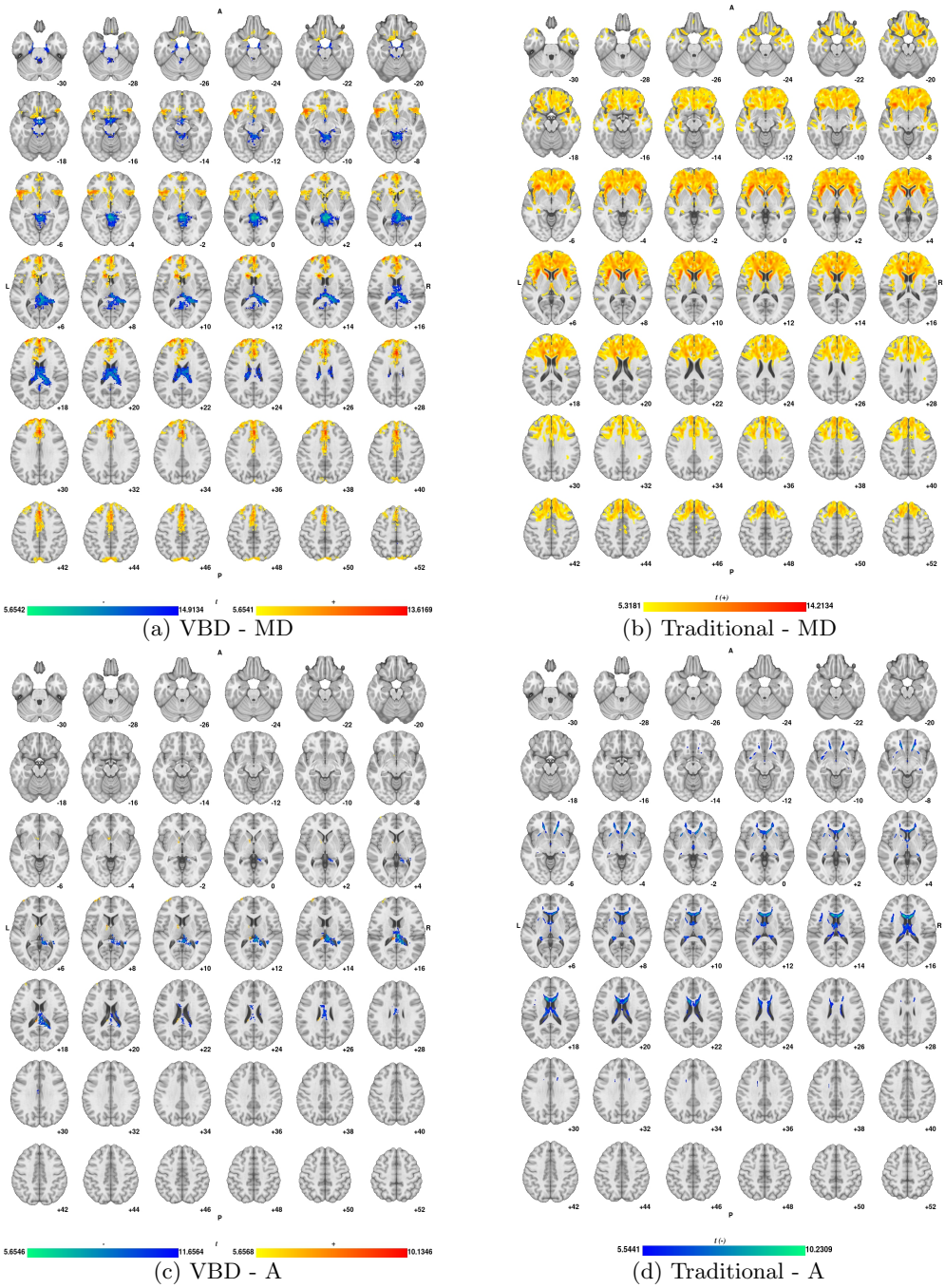
3 Results

3.1 Model Classification Accuracy

Table 2 shows the model performance, reaching a balanced accuracy of 87.8% which is significantly better than chance ($\chi^2 = 113.76$, $p < 10^{-5}$). Although the number of datasets in this problem was relatively small, we were successful in achieving sufficient accuracy not only for biomarker discovery, but potential diagnostic use having a similarly low p-value as the diagnostic questionnaires in Table 1.

3.2 Voxel-Based Diktiometry Maps

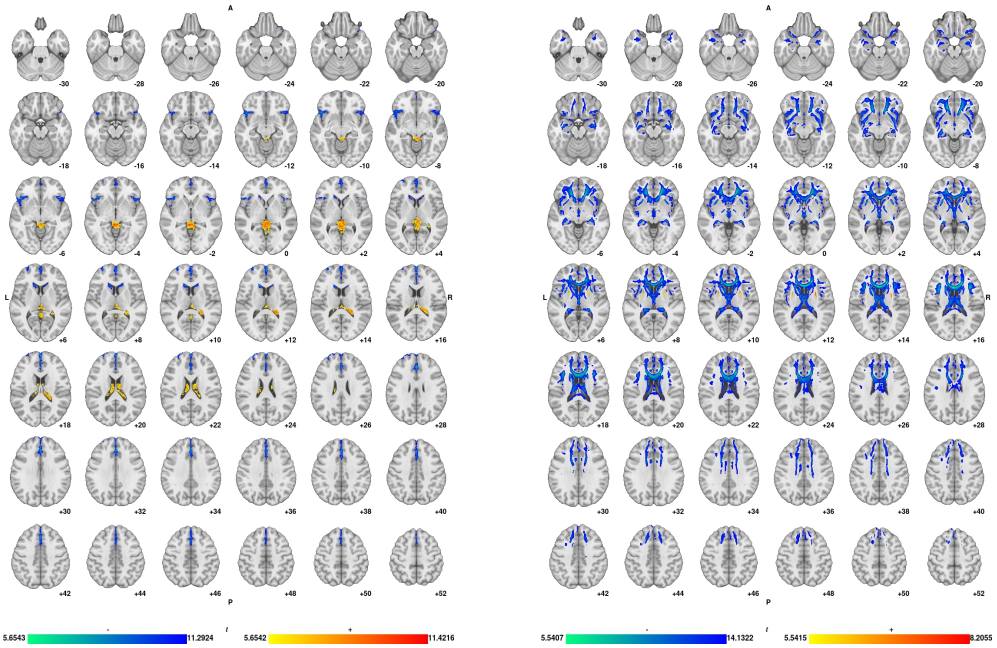
Traditional VBA applied to the same diffusivity measures showed increased MD and reduced FA symmetrically in the frontal lobes which is consistent with the white matter degeneration in bvFTD [24, 25]. Interestingly, VBD analysis shows that the CNNs developed for detecting bvFTD do not depend on these features but find others to be more salient instead. As shown in Figure 2, VBD finds results largely localised to the Sylvian fissure and frontal lobe, likely reflecting the residual or subvoxel morphological changes in this region that are not fully corrected by deformable registration.



(a) VBD - MD (b) Traditional - MD

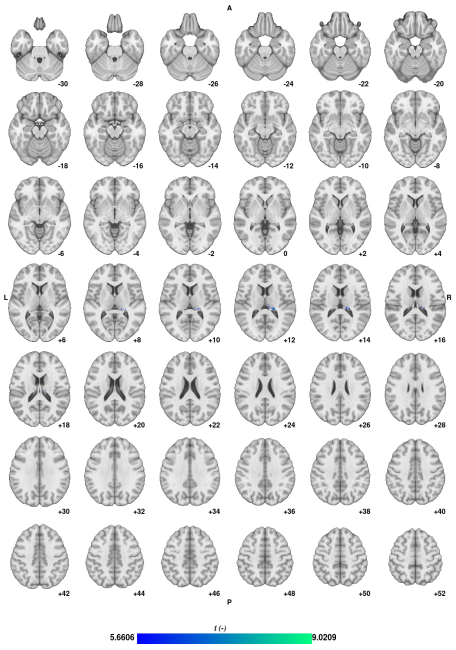
(c) VBD - A (d) Traditional - A

Fig. 2: Voxel-based diktiometry sensitivity results and traditional voxel-based analysis results using the combined dataset (in the first and second columns, respectively) from $z = -30$ to $z = +52$ in MNI space for the mean diffusivity (MD, first row) and anisotropy (A, second row).

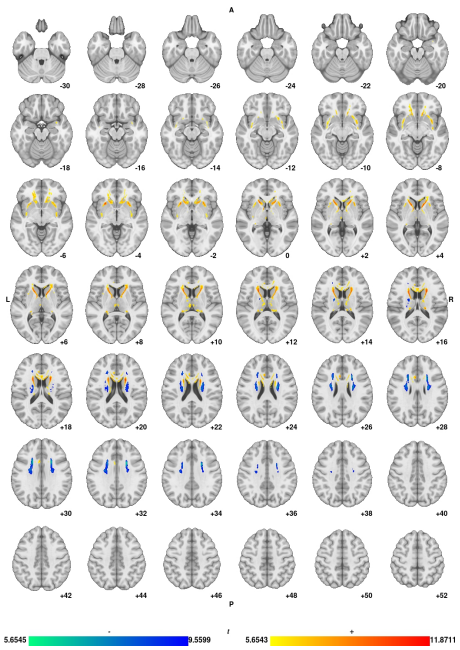


(a) VBD - FA

(b) Traditional - FA



(c) VBD - PsPl



(d) Traditional - PsPl

Fig. 3: Voxel-based diktiometry sensitivity results and traditional voxel-based analysis results using the combined dataset (in the first and second columns, respectively) from $z = -30$ to $z = +52$ in MNI space for the fractional anisotropy (FA, first row) pseudo-planarity (PsPl, second row).

For VBD, the highest significance results are largely constrained to MD in the frontal lobe, temporal lobe, and Sylvian fissure. This is largely consistent with cortical degeneration in these regions. In addition, there is some evidence that they also occur in the parietal lobe as well but to a lesser degree, indicating some amount of widerspread cortical atrophy indicative of various dementia-like disorders [26].

3.3 Sparse principal components maps

sPCA decomposition (Figure 4) revealed four primary axes along which the CNN sensitivities vary within the patient population . These clusters show how the VBD maps can be interpreted as four different axes in terms of how the biomarkers are expressed. Each component is accompanied by scatterplots showing significant correlations ($p < 0.05$) with the clinical scores, summarised in Figure 5.

PC1

The signal for the first decoupled biomarker is concentrated on the mean-diffusivity sensitivity in the frontal cortex proximal to the left Sylvian fissure in the superior temporal gyrus which also was the highest signal location for the VBD maps as a whole. It includes some isolated clusters in the prefrontal cortex as well, more specifically in the medial superior frontal gyrus, middle cingulate and paracingulate gyri (MCG), and middle frontal gyrus (MFG). Some other clusters were located on the amygdala region on the right hemisphere, the pulvinar on left hemisphere and the hippocampus on both hemispheres as well as on the precentral and postcentral gyri (PreCG and PoCG), the precuneus, the superior parietal gyrus (SPG) and the supplementary motor area (SMA). The expression of this particular biomarker was correlated with overall dementia severity via MMSE, two measures of apathy in the ECOCAPTURE dataset, and language in the FTLDNI. Thus, this biomarker appears to capture general bvFTD severity.

PC2

The signal for the second decoupled biomarker is concentrated on the mean-diffusivity and anisotropy sensitivities in the pre-frontal cortex, including the medial SFG, anterior cingulate cortex pregenual (ACCpre), the SFG medial orbital and the inferior frontal gyrus (IFG) triangular part. It is correlated with motor disturbance as well as social disinhibition, both characteristic symptoms of bvFTD, and thus seems to be sensitive to symptoms specifically involving the executive capacities of said region both in terms of social and physical behaviour.

PC3

The third decoupled biomarker is in a similar location as PC1 in terms of being in the frontal lobe proximal to the Sylvian fissure, but somewhat more lateral including the IFG pars orbitals, IFG opercular and IFG triangular parts as well as scattered clusters throughout the pre-frontal lobe, including the MFG, the SMA, the MCG, the anterior cingulate and paracingulate gyri (ACG), PreCG and PoCG on the left

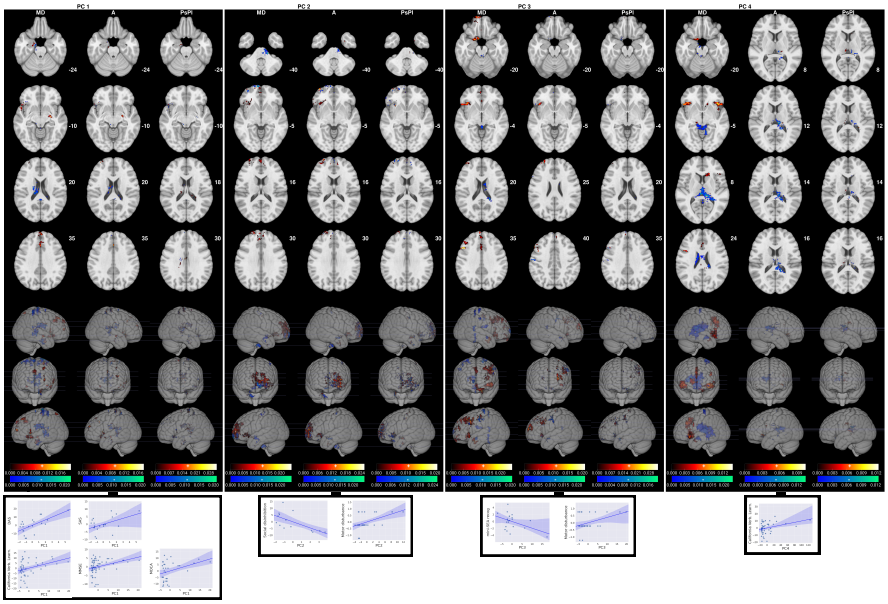


Fig. 4: Visualization of the decoupled biomarkers in MNI space obtained by sPCA on the sensitivity images relative to MD, A, and PsPl. Each column shows the loading contribution to the diffusivity measure limited to clusters of 256 mm^3. The fitting lines correlating the residuals of the eigenvector and the clinical scores ($p < 0.05$) are shown in the scatter plots.

		PC1	PC2	PC3	PC4
ECOCAPTURE					
Apathy	DAS	(0.6415, 0.0017)			
	SAS	(0.4763, 0.029)			
Behavioural metrics	Social disinhibition		(-0.6461, 0.0125)		
Social and Emotions	miniSEA (recognition)			(-0.444, 0.0438)	
FTLDNI					
Cognitive	MOCA	(0.3683, 0.0349)			
Language	California verbal learning	(0.3424, 0.0198)			(0.3329, 0.0238)
Neuropsychiatric Inventory	Motor Disturbance		(0.5027, 0.0039)	(0.4119, 0.0213)	
FTLDNI + ECOCAPTURE					
Cognitive	MMSE	(0.3117, 0.0102)			

Fig. 5: Significant correlations ($p < 0.05$) between decoupled biomarkers and clinical scores (r, p-value).

hemisphere with some clusters occurring also around the thalamus, the amygdala and parahippocampal gyrus (PHG). In terms of correlations, this biomarker was correlated with motor disturbances but less strongly so than PC2. It was more strongly correlated with emotion recognition which might be explained by its spatial proximity to PC1 which also has an affective component.

PC4

The final decoupled biomarker has three primary clusters: one in the third ventricle and the other two in the frontal cortex proximal to the Sylvian fissure similar to PC3 but bilaterally. The clusters were located more specifically in regions of the IFG pars orbitals and triangular parts in both hemispheres as well as in the SFG medial and anterior cingulate cortex supracallosal, MCG, SMA and caudate nucleus. In combination with PC3, this indicates that there is some asymmetric changes in these regions which is more extreme on the left side. The symmetric component of this change is captured by PC4 and the residual left part is captured by PC3. The patient's expression of the PC4 decoupled biomarker was weakly correlated with the patient's score on the verbal learning test, meaning that it might be specific to language.

4 Discussion and Limitations

Both VBA and VBD found a correlation between FTD and increased diffusivity in the frontal and temporal lobes (Figure 2(a)), which is consistent with cortical degeneration, confirming that the primary mechanism of FTD is visible in DTI.

The ability to decompose the VBD results into multiple distinct axis demonstrates the heterogeneity of this cortical degeneration even within a patient cohort with the same FTD subtype. This addresses a concern raised by [8] that many methods in the literature are limited to the comparison between healthy controls and a particular FTD variant which limits their clinical utility.

The correlation between sPCA results and clinical information suggest that even within the behavioural variant of FTD, there still exists some heterogeneity. Certain decoupled biomarkers were more heavily associated with the behavioural aspects of the disorder (notably PC1) whereas others were more heavily related to the cognitive and language aspects (notably PC4). This confirms that the variants are not strictly delineated but co-exist and that multiple symptomatologically-similar neurological disorders could be present in the datasets, such as corticobasal degeneration, noting that our method did find significant sensitivities in the midbrain and that FTD is often seen in patients who develop corticobasal degeneration [27].

Despite being composed of two separate databases, our dataset is also relatively small compared to the extensive medical imaging databases seen for similar domains such as Alzheimer's disease. This limits network size and complexity to traditionally-designed CNNs with fewer of weights and thus lower data requirements.

4.1 Future work

Interpretation is still an active area of research for machine-learning-based biomarker discovery. Some of these interpretational difficulties have been addressed through

sPCA decoupling the sensitivity maps in order to indicate when regions are or are not a part of singular biomarker. However the issue of interpretation of each of these decoupled biomarkers remains as one must still posit the underlying mechanism for the sensitivities seen and why disparate regions may be correlated. Despite this, VBD combined with sPCA still offers a powerful tool for hypothesis generation that would allow for more targeted, specific investigations of particular biomarkers to be performed.

One immediate area of future work would be to extend the analysis to use a larger ensemble of CNNs as well as extensive data augmentation for the classification and sensitivity calculation tasks. Previous studies using VBD have shown that the statistical smoothing resulting from larger ensembles tends to reduce noise [10, 11].

As suggested by our two aforementioned limitations, our most immediate future work is to extend this analysis to a cohort of patients with multiple variants of FTD in order to determine if specific imaging biomarkers can be extracted for each variant. This may confirm that the variants do represent different anatomically separable axes as was the case with the behavioural variant, which could be diagnostically useful for detecting them prior to their characteristic symptom onset.

5 Conclusion

Frontotemporal lobe dementia (FTD) and specifically its behavioural variant, displays very heterogeneous etiology with different patients experiencing different levels and types of symptoms. The goal of this paper is to use machine learning to automatically extract imaging biomarkers for behavioural variant FTD and to develop a framework that can automatically partition these biomarkers into ones that reflect different aspects of this disorder. In order to achieve this, an ensemble of convolution neural networks are trained to classify DTI images between patients and healthy controls, and voxel-based diktiometry is used to extract biomarkers for this task from these networks. sPCA shows that there are at least six axes of heterogeneity in the patient dataset visible in DTI. Four of these axes correlate with clinical symptoms indicates that these distinct biomarkers reflect different symptomatological axes in bvFTD.

Declarations

- **Funding:** Alfonso Estudillo Romero is supported through the SAD Région Bretagne programme and the Institut des Neurosciences Cliniques de Rennes (INCR).

- **Conflict of interest:** The authors have no conflicts of interest to declare.
- **Availability of data and materials:** Data from the ECOCAPTURE dataset [12], [13] was collected by the Insitut du Cerveau (Paris, France). NIFD data is available at https://ida.loni.usc.edu, https://memory.ucsf.edu/research-trials/research/allftd
- **Ethics approval:** All ECOCAPTURE data was collected with institutional ethics board approval (Clinicaltrials.gov:NTC02496312, NCT03272230).

References

[1] Bang, J., Spina, S., Miller, B.L.: Frontotemporal dementia. The Lancet **386**(10004), 1672–1682 (2015)

[2] Laforce Jr, R.: Behavioral and language variants of frontotemporal dementia: a review of key symptoms. Clinical Neurology and Neurosurgery **115**(12), 2405–2410 (2013)

[3] McCarthy, J., Collins, D.L., Ducharme, S.: Morphometric mri as a diagnostic biomarker of frontotemporal dementia: a systematic review to determine clinical applicability. NeuroImage: Clinical **20**, 685–696 (2018)

[4] Perry, R.J., Graham, A., Williams, G., Rosen, H., Erzinçlioglu, S., Weiner, M., Miller, B., Hodges, J.: Patterns of frontal lobe atrophy in frontotemporal dementia: a volumetric mri study. Dementia and geriatric cognitive disorders **22**(4), 278–287 (2006)

[5] Peelle, J.E., Troiani, V., Gee, J., Moore, P., McMillan, C., Vesely, L., Grossman, M.: Sentence comprehension and voxel-based morphometry in progressive nonfluent aphasia, semantic dementia, and nonaphasic frontotemporal dementia. Journal of Neurolinguistics **21**(5), 418–432 (2008)

[6] Bouts, M.J.R.J., Möller, C., Hafkemeijer, A., Swieten, J.C., Dopper, E., Flier, W.M., Vrenken, H., Wink, A.M., Pijnenburg, Y.A.L., Scheltens, P., Barkhof, F., Schouten, T.M., Vos, F., Feis, R.A., Grond, J., Rooij, M., Rombouts, S.A.R.B.: Single Subject Classification of Alzheimer's Disease and Behavioral Variant Frontotemporal Dementia Using Anatomical, Diffusion Tensor, and Resting-State Functional Magnetic Resonance Imaging. Journal of Alzheimer's Disease **62**(4), 1827–1839 (2018)

[7] Feis, R.A., Bouts, M.J.R.J., Panman, J.L., Jiskoot, L.C., Dopper, E.G.P., Schouten, T.M., Vos, F., Grond, J., Swieten, J.C., Rombouts, S.A.R.B.: Single-subject classification of presymptomatic frontotemporal dementia mutation carriers using multimodal MRI. NeuroImage: Clinical **20**, 188–196 (2018)

[8] Pellegrini, E., Ballerini, L., Hernandez, M.d.C.V., Chappell, F.M., González-Castro, V., Anblagan, D., Danso, S., Muñoz-Maniega, S., Job, D., Pernet, C., *et al.*: Machine learning of neuroimaging for assisted diagnosis of cognitive impairment and dementia: a systematic review. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring **10**, 519–535 (2018)

[9] Ahmed, M.R., Zhang, Y., Feng, Z., Lo, B., Inan, O.T., Liao, H.: Neuroimaging and machine learning for dementia diagnosis: recent advancements and future prospects. IEEE reviews in biomedical engineering **12**, 19–33 (2018)

[10] Estudillo-Romero, A., Haegelen, C., Jannin, P., Baxter, J.S.H.: Voxel-based diktiometry: Combining convolutional neural networks with voxel-based analysis and its application in diffusion tensor imaging for Parkinson's disease. Human Brain Mapping, 1–17 (2022)

[11] Estudillo-Romero, A., Migliaccio, R., Batrancourt, B., Jannin, P., Baxter, J.S.:

Non-local diffusion-based biomarkers in patients with cocaine use disorder. Neuroimage: Reports **4**(2), 100202 (2024)

[12] Tanguy, D., Batrancourt, B., Estudillo-Romero, A., Baxter, J.S.H., Le Ber, I., Bouzigues, A., Godefroy, V., Funkiewiez, A., Chamayou, C., Volle, E., Saracino, D., Rametti-Lacroux, A., Morandi, X., Jannin, P., Levy, R., Migliaccio, R.: An ecological approach to identify distinct neural correlates of disinhibition in frontotemporal dementia. NeuroImage: Clinical **35**, 103079 (2022)

[13] Godefroy, V., Levy, R., Bouzigues, A., Rametti-Lacroux, A., Migliaccio, R., Batrancourt, B.: Ecocapture@ home: Protocol for the remote assessment of apathy and its everyday-life consequences. International journal of environmental research and public health **18**(15), 7824 (2021)

[14] Olney, N.T., Ong, E., Goh, S.M., Bajorek, L., Dever, R., Staffaroni, A.M., Cobigo, Y., Bock, M., Chiang, K., Ljubenkov, P., Kornak, J., Heuer, H.W., Wang, P., Rascovsky, K., Wolf, A., Appleby, B., Bove, J., Bordelon, Y., Brannelly, P., Brushaber, D., Caso, C., Coppola, G., Dickerson, B.C., Dickinson, S., Domoto-Reilly, K., Faber, K., Ferrall, J., Fields, J., Fishman, A., Fong, J., Foroud, T., Forsberg, L.K., Gearhart, D.J., Ghazanfari, B., Ghoshal, N., Goldman, J., Graff-Radford, J., Graff-Radford, N.R., Grant, I., Grossman, M., Haley, D., Hsiung, G., Huey, E.D., Irwin, D.J., Jones, D.T., Kantarci, K., Karydas, A.M., Kaufer, D., Kerwin, D., Knopman, D.S., Kramer, J.H., Kraft, R., Kremers, W., Kukull, W., Lapid, M.I., Litvan, I., Mackenzie, I.R., Maldonado, M., Manoochehri, M., McGinnis, S.M., McKinley, E.C., Mendez, M.F., Miller, B.L., Onyike, C., Pantelyat, A., Pearlman, R., Petrucelli, L., Potter, M., Rademakers, R., Ramos, E.M., Rankin, K.P., Roberson, E.D., Rogalski, E., Sengdy, P., Shaw, L.M., Syrjanen, J., Tartaglia, M.C., Tatton, N., Taylor, J., Toga, A., Trojanowski, J.Q., Weintraub, S., Wong, B., Wszolek, Z., Boxer, A.L., Boeve, B.F., Rosen, H.J., on behalf of the ARTFL and LEFFTDS consortia: Clinical and volumetric changes with increasing functional impairment in familial frontotemporal lobar degeneration. Alzheimer's & Dementia **16**(1), 49–59 (2020)

[15] Descoteaux, M., Wiest-Daesslé, N., Prima, S., Barillot, C., Deriche, R.: Impact of rician adapted non-local means filtering on hardi. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008, pp. 122–130. Springer, Berlin, Heidelberg (2008)

[16] Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4itk: Improved n3 bias correction. IEEE Transactions on Medical Imaging **29**(6), 1310–1320 (2010)

[17] Tournier, J.-D., Smith, R., Raffelt, D., Tabbara, R., Dhollander, T., Pietsch, M., Christiaens, D., Jeurissen, B., Yeh, C.-H., Connelly, A.: MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation.

NeuroImage **202**, 116137 (2019)

[18] Andersson, J.L.R., Sotiropoulos, S.N.: An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. NeuroImage **125**(C), 1063–1078 (2016)

[19] Norton, I., Essayed, W.I., Zhang, F., Pujol, S., Yarmarkovich, A., Golby, A.J., Kindlmann, G., Wassermann, D., Estepar, R.S.J., Rathi, Y., Pieper, S., Kikinis, R., Johnson, H.J., Westin, C.-F., O'Donnell, L.J.: SlicerDMRI: Open Source Diffusion MRI Software for Brain Cancer Research. Cancer Research **77**(21), 101–103 (2017)

[20] Avants, B., Epstein, C., Grossman, M., Gee, J.: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis **12**(1), 26–41 (2008)

[21] Avants, B.B., Tustison, N.J., Song, G., Cook, P.A., Klein, A., Gee, J.C.: A reproducible evaluation of ANTs similarity metric performance in brain image registration. NeuroImage **54**(3), 2033–2044 (2011)

[22] Fonov, V., Evans, A., McKinstry, R., Almli, C., Collins, D.: Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. NeuroImage **47**, 102 (2009). Organization for Human Brain Mapping 2009 Annual Meeting

[23] Zou, H., Hastie, T., Tibshirani, R.: Sparse principal component analysis. Journal of computational and graphical statistics **15**(2), 265–286 (2006)

[24] Agosta, F., Scola, E., Canu, E., Marcone, A., Magnani, G., Sarro, L., Copetti, M., Caso, F., Cerami, C., Comi, G., *et al.*: White matter damage in frontotemporal lobar degeneration spectrum. Cerebral cortex **22**(12), 2705–2714 (2012)

[25] Lam, B.Y., Halliday, G.M., Irish, M., Hodges, J.R., Piguet, O.: Longitudinal white matter changes in frontotemporal dementia subtypes. Human Brain Mapping **35**(7), 3547–3557 (2014)

[26] Rajagopalan, V., Pioro, E.P.: Distinct patterns of cortical atrophy in als patients with or without dementia: an mri vbm study. Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration **15**(3-4), 216–225 (2014)

[27] Lee, S.E., Rabinovici, G.D., Mayo, M.C., Wilson, S.M., Seeley, W.W., DeArmond, S.J., Huang, E.J., Trojanowski, J.Q., Growdon, M.E., Jang, J.Y., *et al.*: Clinicopathological correlations in corticobasal degeneration. Annals of neurology **70**(2), 327–340 (2011)