NeuroImage: Clinical 37 (2023) 103320

Contents lists available at ScienceDirect

NeuroImage: Clinical

journal homepage: www.elsevier.com/locate/ynicl

Multiclass prediction of different dementia syndromes based on multi-centric volumetric MRI imaging

Leonie Lampe [a,b], Hans-Jürgen Huppertz [c], Sarah Anderl-Straub [d], Franziska Albrecht [a], Tommaso Ballarini [a], Sandrine Bisenius [a], Karsten Mueller [a], Sebastian Niehaus [a,o], Klaus Fassbender [e], Klaus Fliessbach [f], Holger Jahn [g], Johannes Kornhuber [h], Martin Lauer [i], Johannes Prudlo [j], Anja Schneider [f,k], Matthis Synofzik [l], Jan Kassubek [d], Adrian Danek [m], Arno Villringer [a,b], Janine Diehl-Schmid [n], Markus Otto [d,p], Matthias L. Schroeter [a,b,*], FTLD Consortium Germany

[a] *Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*
[b] *Clinic for Cognitive Neurology, University Clinic Leipzig, Germany*
[c] *Swiss Epilepsy Clinic, Klinik Lengg, Zurich, Switzerland*
[d] *Department of Neurology, University of Ulm, Germany*
[o] *Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany*
[e] *Department of Neurology, Saarland University, Homburg, Germany*
[f] *Clinic for Neurodegenerative Diseases and Geriatric Psychiatry, University of Bonn, and German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany*
[g] *Clinic for Psychiatry and Psychotherapy, University Hospital Hamburg-Eppendorf, Germany*
[h] *Department of Psychiatry and Psychotherapy, Friedrich-Alexander-University of Erlangen-Nuremberg, Erlangen, Germany*
[i] *Department of Psychiatry and Psychotherapy, University Wuerzburg, Germany*
[j] *Department of Neurology, University of Rostock, and DZNE, Rostock, Germany*
[k] *Department of Psychiatry and Psychotherapy, University of Goettingen, Germany*
[l] *Department of Neurodegenerative Diseases, Centre for Neurology & Hertie-Institute for Clinical Brain Research, University of Tuebingen, Germany & DZNE, Tuebingen, Germany*
[m] *Department of Neurology, Ludwig-Maximilians-Universität Munich, München, Germany*
[n] *Department of Psychiatry and Psychotherapy, Technical University of Munich, Germany*
[p] *Department of Neurology, University of Halle, Germany*

A R T I C L E I N F O

Keywords:
Dementia
Diagnosis
Machine learning
MRI
Neurodegeneration
Volumetry

A B S T R A C T

Introduction: Dementia syndromes can be difficult to diagnose. We aimed at building a classifier for multiple dementia syndromes using magnetic resonance imaging (MRI).
Methods: Atlas-based volumetry was performed on T1-weighted MRI data of 426 patients and 51 controls from the multi-centric German Research Consortium of Frontotemporal Lobar Degeneration including patients with behavioral variant frontotemporal dementia, Alzheimer's disease, the three subtypes of primary progressive aphasia, i.e., semantic, logopenic and nonfluent-agrammatic variant, and the atypical parkinsonian syndromes progressive supranuclear palsy and corticobasal syndrome. Support vector machine classification was used to classify each patient group against controls (binary classification) and all seven diagnostic groups against each other in a multi-syndrome classifier (multiclass classification).
Results: The binary classification models reached high prediction accuracies between 71 and 95% with a chance level of 50%. Feature importance reflected disease-specific atrophy patterns. The multi-syndrome model reached accuracies of more than three times higher than chance level but was far from 100%. Multi-syndrome model performance was not homogenous across dementia syndromes, with better performance in syndromes characterized by regionally specific atrophy patterns. Whereas diseases generally could be classified vs controls more correctly with increasing severity and duration, differentiation between diseases was optimal in disease-specific windows of severity and duration.

* Corresponding author at: Max Planck Institute for Human Cognitive and Brain Sciences, & Clinic for Cognitive Neurology, University Clinic Leipzig, Germany.
E-mail address: schroet@cbs.mpg.de (M.L. Schroeter).

https://doi.org/10.1016/j.nicl.2023.103320
Received 11 August 2022; Received in revised form 23 November 2022; Accepted 4 January 2023
Available online 5 January 2023

Discussion: Results suggest that automated methods applied to MR imaging data can support physicians in diagnosis of dementia syndromes. It is particularly relevant for orphan diseases beside frequent syndromes such as Alzheimer's disease.

1. Introduction

Dementia syndromes represent a major burden for aging societies around the globe. They are related to the neurodegenerative disease spectrum, which includes various forms of dementia. This study specifically focuses on Alzheimer's disease (AD) (Warren et al., 2012; McKhann et al., 2011), frontotemporal lobar degeneration (FTLD) with behavioral variant frontotemporal dementia (bvFTD) (Rascovsky et al., 2011), and primary progressive aphasias (PPA) reflected as three subforms, i.e., semantic variant (svPPA), nonfluent-agrammatic variant (nfvPPA), and logopenic variant (lvPPA) (Gorno-Tempini et al., 2011), atypical parkinsonian syndromes also associated with dementia such as corticobasal syndrome (CBS) and progressive supranuclear palsy (PSP) (Meeter et al., 2017). Neurodegenerative diseases primarily affect several functions from memory (AD) to behavior/personality (bvFTD), language (PPAs), and motor functions (CBS, PSP).

Recently, imaging as well as molecular biomarkers have been incorporated into the diagnostic criteria for AD, bvFTD, and PPAs (Gorno-Tempini et al., 2011; McKhann et al., 2011; Rascovsky et al., 2011). Comprehensive quantitative and systematic meta-analyses validated and confirmed disease-specific atrophy patterns in the aforementioned neurodegenerative diseases (Schroeter et al., 2007; Schroeter et al., 2008; Schroeter et al., 2009; Schroeter et al., 2014; Schroeter and Neumann, 2011; Bisenius et al., 2016; Albrecht et al., 2017; Whitwell et al., 2017). While this conceptual change aids to diagnose distinct syndromes more accurately, imaging biomarkers specifically can be a challenge for the radiologist lacking specific training (Klöppel et al., 2008). The atrophy patterns of different dementia syndromes are complex and are easily overseen by non-specialists (Schroeter et al., 2007; Schroeter et al., 2008; Schroeter et al., 2009; Schroeter et al., 2014; Schroeter and Neumann, 2011; Bisenius et al., 2016; Albrecht et al., 2017; Whitwell et al., 2017). Less common 'orphan' dementia syndromes can lead to misdiagnosis. Clinicians use imaging results and embed them into further information from clinical history, neuropsychological testing, and other biomarkers to diagnose disease and initiate adequate treatment (Gorno-Tempini et al., 2011; McKhann et al., 2011; Rascovsky et al., 2011). Neglecting differential diagnosis can, on the one hand, hamper selective treatment, and on the other hand, preclude the individual and their families from obtaining specialized support. In times of low density of expertise and medical care in rural regions, aging populations are confronted with higher rates of misdiagnosis and limited capabilities to provide appropriately tailored treatment for patients (Goins et al., 2005).

Computerized methods involving artificial intelligence have proven to be well suited for complex multivariate data and could be a valuable addition in the diagnostic process by providing decision support (Dwyer et al., 2018). In fact, machine learning approaches for pattern recognition have already been successfully applied to consistently detect skin cancer (Esteva et al., 2017), lung cancer (Yu et al., 2016), and eye diseases (De Fauw et al., 2018) or predict cognitive decline from multimodal imaging data in AD (Franzmeier et al., 2020). Since some of these automated methods even out-performed specialists (Klöppel et al., 2008), they could be a valuable addition to support physicians in finding the right differential diagnosis as early as possible in the disease course.

Recently, our lab has developed such approaches to disentangle diagnosis and differential diagnosis and predict therapeutic efficacy for several of these neurodegenerative diseases based on multimodal imaging (Dukart et al., 2011; Meyer et al., 2017; Mueller et al., 2017; Bisenius et al., 2017; Anderl et al., 2021; Koutsouleris et al., 2022). To pave the road for translation of these approaches into clinical routine, classifiers containing multiple syndromes should be explored as feasible options (Arbabshirani et al., 2017). While binary classifiers separating healthy individuals from demented patients have yielded high performance measures (Magnin et al., 2009; Arbabshirani et al., 2017), the next step is to include multiple dementia syndromes to differentiate them from each other (Lampe et al., 2022). This differential diagnostic stance has to be regarded as the most challenging and most appropriate if translation to clinical routine is the primary aim. Recently, Huppertz and colleagues (2016) have demonstrated the feasibility of such an approach in distinguishing neurodegenerative parkinsonian syndromes. Other studies have compared two dementia disease groups against each other, such as AD vs bvFTD (Zheng et al., 2016; Tahmasian et al., 2016) or two out of three PPA subtypes (Bisenius et al., 2017).

Here, we explore the application of support vector machines (SVM), a multivariate supervised learning algorithm, on structural volumetric magnetic resonance imaging (MRI) data to differentiate multiple dementia syndromes first from healthy individuals, and secondly from multiple other dementia syndromes in a differential diagnostic approach. We hypothesize that both binary classifiers as well as the multi-syndrome classifiers can reach high accuracies in differentiating syndromes based on the assumption of disease-specific atrophy fingerprints. This assumption is supported by recent meta-analyses/reviews revealing high classification accuracies for neurodegenerative diseases in contrast to psychiatric disorders (Arbabshirani et al., 2017; Woo et al., 2017). We focus on structural MRI as it is the most common and widely available method.

2. Methods

2.1. Participants

Our study included a large multicenter cohort with 426 patients and 51 healthy controls from the German Research Consortium of FTLD (Otto et al., 2011). The patient cohort consisted of 72 patients with AD, 146 patients with bvFTD, 26 patients with CBS, 30 patients with lvPPA, 58 patients with nfvPPA, 48 patients with PSP, and 46 patients with svPPA. All patients were thoroughly examined clinically according to standardized operating procedures (SOPs) and were classified based on established international diagnostic criteria (Otto et al., 2011). Diagnosis was primarily based on clinical criteria, supported by imaging (structural MRI, molecular imaging for amyloid / tau) and biomarkers from cerebrospinal fluid (amyloid, tau, phospho tau), i.e., for increasing validity; see for instance for AD, bvFTD, and PPAs (Gorno-Tempini et al., 2011; McKhann et al., 2011; Rascovsky et al., 2011; Warren et al., 2012). Histopathological information was available for single subjects only. Genetic analysis results had been not available for single subjects in our study. However, a recent study (Wagner et al., 2021) involving partly overlapping 509 patients with bvFTD and PPA from the FTLD consortium study revealed by exome sequencing as well as C9orf72 repeat analysis that 18.1 % did show pathogenic variants, without any impact of distribution of APOE alleles. Note that necessary fulfillment of operationalized clinical criteria as a precondition for inclusion in our study prevented from circular approaches regarding imaging. Table 1 gives an overview over age, disease duration, and cognitive test results. We did not collect information about the participants' race or ethnicity. Participants were belonging mainly to central European ethnicity. Differences between distributions were tested with the Kruskal-Wallis test and post hoc with a Wilcoxon rank-sum test between all pairs of samples (Bonferroni-corrected).

Ethical approval

The study was conducted according to the Declaration of Helsinki. It was approved by the local ethics committees of all participating centers (University of Ulm: #39/11, 8th March 2011). Patients, participants, caregivers, or legal representatives gave written informed consent for the study.

2.3. Acquisition of imaging data

Structural MRI head scans were acquired multi-centrically at university hospitals in Bonn (33), Erlangen (40), Goettingen (36), Hamburg (42), Homburg (54), Leipzig (68), Munich – Ludwig-Maximilians-Universität (LMU) (50); Munich – Technische Universität (TU) (137), Rostock (42), Tuebingen (17), Ulm (191) and Wuerzburg (25). Every subject obtained a T1-weighted three-dimensional (1 mm isovoxel resolution) magnetization prepared rapid gradient echo (MPRAGE) head MRI brain scan (Brant-Zawadzki et al., 1992). The MPRAGE sequences were converted to ANALYZE 7.5 format.

2.4. Analysis of imaging data

Atlas-based volumetry (ABV) was applied to the MPRAGE sequence data to determine volumes or areas of 64 brain regions. ABV classifies the image on voxel level into gray matter, white matter, and cerebrospinal fluid compartments and warps the resulting tissue probability maps into a template space using elastic image registration. Subsequently, it employs an atlas of predefined regions of interest in the same space to extract regional brain volumes or areas. For this study, the LONI Probabilistic Brain Atlas (LPBA40) (Shattuck et al., 2008) and further masks derived from it were utilized. A detailed description of all image processing steps and atlas regions included can be found in Huppertz et al. (2016). Before being used as prediction features, all ABV results were normalized to total intracranial volume.

For SVM classification, we used the freely available e1071 package implemented in the caret package in the R environment (Vapnik, 1995). SVM is based on libsvm, an open-source machine learning library (Vapnik, 1995). In SVM classification, optimal separating hyperplanes are defined, which maximize the distance between subjects belonging to different classes (here syndromes). These hyperplanes are set in a virtual vector space, resulting in a non-linear classification function. This eliminates the need for static tests such as testing the normal distribution (Schölkopf et al., 1999). Classifiers were trained on the ABV results of the 64 brain regions listed in Fig. 1. During training the classifiers, feature importance for classification was determined for every brain region as done in similar studies (see for instance Klöppel et al., 2008; Bron et al., 2015; cf. also "How could I generate the primal variable w of linear SVM?" in https://www.csie.ntu.edu.tw/~cjlin/libsvm/faq.html#f804). A cross-validation of the method was performed by using the leave-one-(subject)-out method (so called leave-one-out cross-validation, LOOCV). This procedure iteratively leaves out the information of one subject at a time and trains the SVM classifier on the remaining data for subsequent class assignment of the left-out subject respectively. This validation allows for the generalization of the trained SVM classifier to unknown data and mitigates the possibility of falsely inflated accuracies. A linear kernel was used for each SVM, and the tuning 'cost' parameter

Table 1
Clinical and demographic characteristics for patients and controls.

Diagnosis	n	Age in years	Disease duration in years	CDR	FTLD-CDR	MMSE
AD	72	66.67 ± 9.59 69.0 (49.0; 33–82) (8.33 %)	3.1 ± 2.84 2 (16; −1–15*) (15.27 %)	5.78 ± 3.36 5 (17.5: 0.5–18) (13.89 %)	6.98 ± 4.22 6 (23.5; 0.5–24) (15.27 %)	21.29 ± 5.86 22.5 (30; 0–30) (8.33 %)
bvFTD	146	61.68 ± 9.67 61 (50; 31–81) (0.68 %)	3.77 ± 4.55 2 (25; −1–24*) (3.42 %)	5.74 ± 3.97 5 (17.5; 0.5–18) (9.59 %)	7.87 ± 5.13 6.5 (23.5; 0.5–24) (9.59 %)	24.63 ± 4.96 26 (30; 0–30) (8.22 %)
CBS	26	65.96 ± 6.91 67 (24; 55–79)	3.17 ± 2.57 2 (9; 0–9) (11.54 %)	4.88 ± 3.46 4 (13; 0–13) (23.08 %)	6.05 ± 4.07 5.75 (15.5; 0.5–16) (23.08 %)	22.81 ± 5.37 23 (21; 9–30) (19.23 %)
lvPPA	30	67.33 ± 5.60 68 (18; 56–74)	3.89 ± 4.25 2 (17; 0–17) (3.33 %)	3.28 ± 2.94 2.5 (10; 0–10) (10 %)	5.31 ± 3.58 4.5 (12.5; 1–13.5) (10 %)	21 ± 5.45 21.5 (18; 10–28) (3.33 %)
nfvPPA	58	68.46 ± 8.32 70 (36; 44–80) (1.72 %)	2.14 ± 1.54 2 (8; 0 – 8) (3.45 %)	2.40 ± 2.47 2 (13; 0–13) (15.52 %)	4.63 ± 3.11 4 (17.5; 0.5–18) (15.52 %)	23.22 ± 6.09 25 (23; 7–30) (3.45 %)
PSP	48	69.06 ± 7.31 68 (34; 50–84)	3.02 ± 1.87 2.75 (16; 0–8) (8.33 %)	4.28 ± 3.88 2.75 (16; 0–16) (33.33 %)	5.61 ± 4.87 4 (20.5; 0.5–21) (35.42 %)	25.38 ± 3.96 27 (16; 14–30) (6.25 %)
svPPA	46	62.14 ± 8.31 61.0 (31; 44–75) (4.34 %)	3 ± 2.08 2.5 (10; 0–10) (4.34 %)	4.02 ± 3.13 3.5 (15; 0–15) (10.86 %)	6.41 ± 4.15 5.0 (19.5; 1.5–21) (10.86 %)	21.02 ± 7.96 24.0 (30; 0–30) (4.34 %)
Controls	51	64.36 ± 12.7 68 (62; 24–86) (1.96 %)	n.a.	0.06 ± 0.195 0 (1; 0–1) (15.69 %)	0.09 ± 0.294 0 (1.5; 0 – 1.5) (15.69 %)	29 ± 0.881 29 (3; 27–30) (1.96 %)
Statistical differences						
Controls vs patients		n.s.	n.a.	t = −23.96; p < 2.2e−16	t = −26.707; p < 2.2e−16	t = 18.213; p < 2.2e−16
Kruskal-Walis (patients and controls)		χ2 = 44.744; p = 1.533e−7	n.a.	χ2 = 153.97; p < 2.2e−16	χ2 = 136.72; p < 2.2e−16	χ2 = 119.85; p < 2.2e−16
Kruskal-Walis (patients)		χ2 = 45.561; p = 3.621e−8	χ2 = 6.336; p = 0.3867	χ2 = 57.452; p = 1.48e−10	χ2 = 28.351; p = 8.068e−5	χ2 = 37.179; p = 1.625e−6

Note: Data are reported as mean ± standard deviation, median, range with min and max. If values were missing, the percentage of missing values is given in brackets. Abbreviations: AD Alzheimer's disease; bvFTD behavioral variant frontotemporal dementia; CBS corticobasal syndrome; CDR clinical dementia rating scale; FTLD-CDR Frontotemporal lobar degeneration-modified CDR; lvPPA logopenic variant primary progressive aphasia; MMSE mini mental state examination; n.a. not applicable; nfvPPA nonfluent variant primary progressive aphasia; n.s. non-significant; PSP progressive supranuclear palsy; svPPA semantic variant primary progressive aphasia. *Note that for two participants (one with AD, the other with bvFTD) disease duration was −1 year, because diagnostic criteria for respective diseases were completely fulfilled one year after imaging, although partly at date of imaging already.

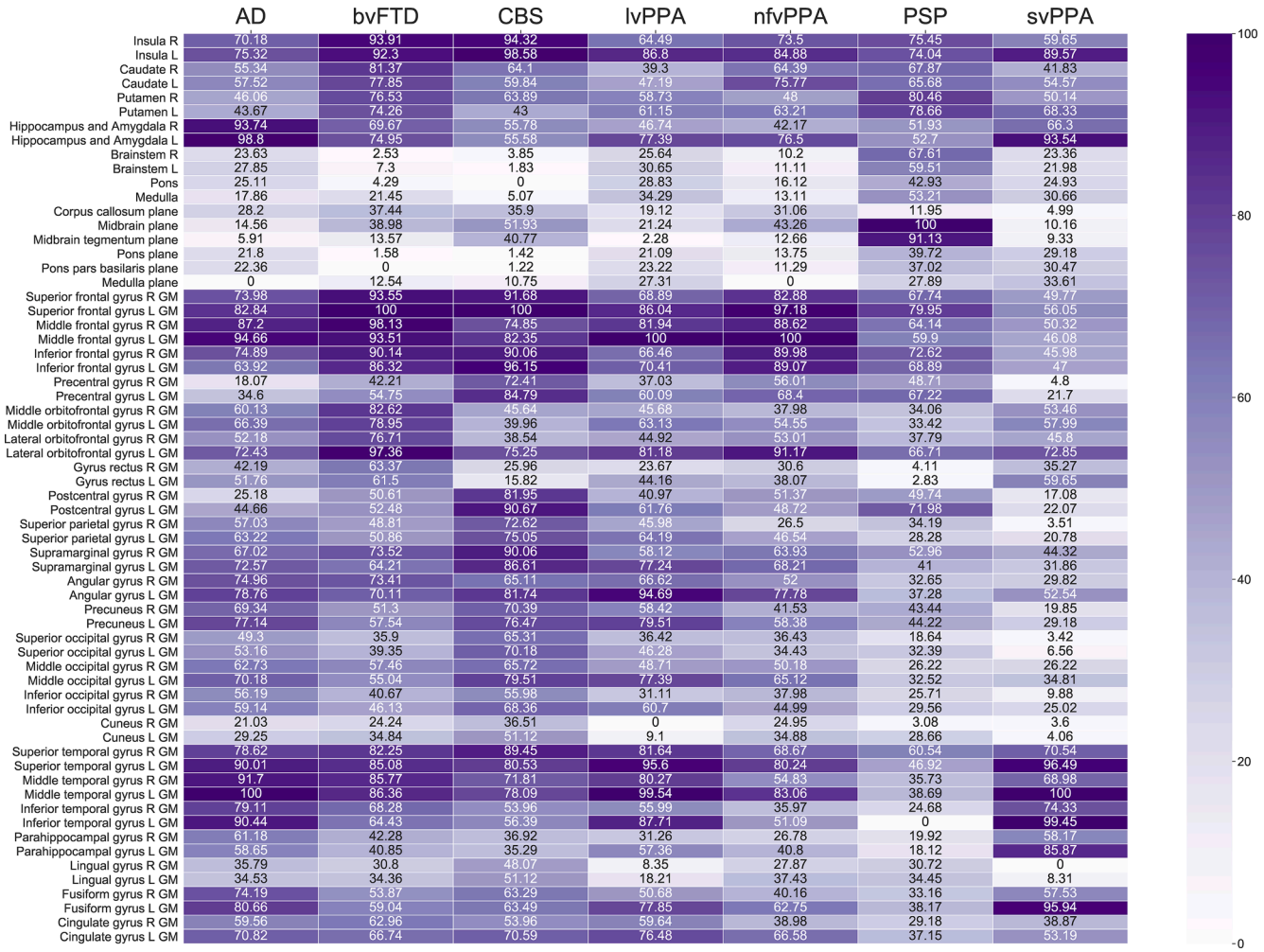

Fig. 1. Importance of brain regions for separating each dementia syndrome from healthy controls. Note: The scale on the right shows the importance for classification from 0 to 100 color coded from white to dark purple. Abbreviations: AD Alzheimer's disease; bvFTD behavioral variant frontotemporal dementia; CBS corticobasal syndrome; GM gray matter; L left; lvPPA logopenic variant primary progressive aphasia; nfvPPA nonfluent variant primary progressive aphasia; PSP progressive supranuclear palsy; R right; svPPA semantic variant primary progressive aphasia.

was held constant at a value of 0.25. Class-probabilities were calculated with Platt scaling as discussed by the authors of LibSVM (Lin et al., 2007).

The SVM is applied using the default parametrization without hyperparameter tuning to avoid overfitting. Furthermore, this setup is reproducible for comparable classification problems and encourages the field to use machine learning algorithms. The diagnostic performance of the binary SVM models (patients with one syndrome vs healthy controls) was assessed by calculating sensitivity, specificity, positive predictive value, negative predictive value, model accuracy, and balanced accuracy (Brodersen et al., 2010). Balanced accuracy is a combined measure of sensitivity and specificity. Additionally, for the binary SVM models the Matthew's correlation coefficient was calculated to assess the classification performance accounting for imbalanced group sizes and chance prediction for binary classifiers (Matthews, 1975).

For multi-class prediction we focused on the sensitivity, the positive predictive value, and the F-score, which is a combined measure of sensitivity and positive predictive value. Note that model metrics for multi-class predictions are often described with precision (which is the positive predictive value here), recall (which is sensitivity here), and the F-score. We further calculated model accuracy for the multi-class SVM.

3. Results

3.1. Descriptive characteristics

Table 1 illustrates clinical and demographic characteristics for patients and controls including age, disease duration, and global neurocognitive scores, i. e., Clinical Dementia Rating (CDR), FTLD-Modified CDR (FTLD-CDR), and Mini Mental State Examination (MMSE) scales. There were significant differences for age for patients with bvFTD being younger than patients with AD ($p < 0.05$), PSP ($p < 0.001$), lvPPA ($p < 0.05$), and nfvPPA ($p < 0.001$). Furthermore, patients with svPPA were significantly younger than patients with PSP ($p < 0.05$) and nfvPPA ($p < 0.05$). Note that there were no significant differences for mean age between controls and the several disease cohorts, excluding age as a bias, at least in the binary syndrome classification. There was no significant difference in disease duration between all patient groups.

As expected, all patient groups had significantly higher CDR scores than healthy controls ($p < 0.001$). Patients with AD had significantly higher CDR scores than patients with lvPPA ($p < 0.05$), nfvPPA ($p < 0.001$), and svPPA ($p < 0.05$). Patients with bvFTD scored significantly worse in the CDR than patients with lvPPA ($p < 0.05$) and nfvPPA ($p < 0.001$). Again, as expected all patient groups had significantly higher scores in the FTLD-CDR than healthy controls ($p < 0.001$). Patients with bvFTD had significantly higher FTLD-CDR scores than patients with

nfvPPA ($p < 0.001$). Additionally, patients with AD had significantly higher FTLD-CDR scores than patients with nfvPPA ($p < 0.05$). All patient groups had significantly lower scores in the MMSE than healthy controls ($p < 0.001$). Patients with AD had significantly lower MMSE scores than patients with PSP ($p < 0.001$) and bvFTD ($p < 0.001$). Also, patients with lvPPA had significantly higher MMSE scores than patients with PSP ($p < 0.05$) and bvFTD ($p < 0.05$). There was no significant difference of gender distribution across patient groups ($\chi^2 = 9.4671$, $p = 0.22$).

3.2. Binary syndrome classification

Results for binary syndrome classification, i.e., identifying dementia in patients vs healthy controls, are illustrated in Table 2. The binary SVM yielded high prediction accuracies between 71 and 95 % (chance level 50 %). To maximize the amount of training data, all patients, and healthy subjects available were used, which might have led to unevenly distributed samples. The Matthews Correlation Coefficient helps to assess the quality of the prediction after accounting for uneven sample sizes. It can be interpreted as [0.70–1.00] very strong prediction, [0.40–0.69] strong prediction, [0.30–0.39] moderate prediction, [0.20–0.29] weak prediction, and [0.01–0.19] negligible relationship. The classification of patients with AD, PSP, and svPPA respectively produced very strong prediction results, while bvFTD, lvPPA, nfvPPA returned strong prediction results. Only CBS generated moderate prediction results. While CBS yielded generally the lowest prediction results, the prediction performance of svPPA was highest (see Table 2) despite the comparatively small sample size of 46 patients.

3.3. Importance of brain regions for binary classification

Different brain regions contribute to the differentiation of the respective dementia syndromes from healthy subjects. Fig. 1 provides an overview over all brain regions that were used in the classifier with their importance for the prediction of each respective syndrome from 0 to 100 (color coded with white to dark purple). Regions overlapped particularly in the frontal and temporal cortex as expected for FTLD subtypes. While brain regions such as the brainstem or the midbrain had low predictive value for most dementia syndromes, they were of high relevance for PSP, in accordance with meta-analyses and the literature (Whitwell et al., 2017). Also in accordance with well-known atrophy patterns, left temporal regions were essential for svPPA, the hippocampus was most crucial for AD and the frontal cortex for bvFTD (Schroeter et al., 2007; Schroeter et al., 2014; Bisenius et al., 2017). As expected, left hemispheric regions were more important than their right hemispheric counterparts for classification of PPAs (Gorno-Tempini et al., 2011; Bisenius et al., 2017). Regions identified as classifier-relevant coincided with atrophic regions identified as disease-specific in meta-analyses and diagnostic criteria in svPPA, lvPPA and nfvPPA (Gorno-Tempini et al., 2011; Bisenius et al., 2017). Note that for CBS several rather regionally unspecific regions were identified as relevant for classification in accordance with recent meta-analyses (Albrecht et al., 2017).

3.4. Multi-syndrome classification

For multi-syndrome classification, all seven dementia syndromes were classified against each other with SVM using LOOCV. Note that, in this sample, no healthy subjects were included because the aim was to simulate the process of radiological differential diagnosis between syndrome-associated atrophy patterns. The sample comprised 426 patients, classified based on 64 predictors (see atlas-based brain regions as described in Fig. 1). Table 3 provides the overview performance measures of the model. The overall accuracy of the multi-syndrome classifier was 47.4 %. Note, that chance level would be here not 50 % like in binary classification, but mathematically 100 % / seven groups, i.e., 14.29 %. Note, however, that this value is given for orientation only as it is modified by several factors such as group size of disease and numbers of respective (disease) controls. Highest performance measures were achieved for classifying svPPA, bvFTD, and PSP, while classification of CBS was weak. For AD, lvPPA, and nfvPPA performance reached reasonable values. Note that specificity, negative predictive value, and balanced accuracy are not informative for multi-class predictions and hence are not given here.

Placeholder Table 3.

3.5. Syndrome probabilities of multi-syndrome classification

Based on the multiclass prediction of different dementia syndromes, probabilities of every patient were calculated for each individual syndrome. The probability values are the transformed output of the SVM. The class division and the subdivision of the vector space of the SVM were transformed into a probability value using a probability calibration method. Fig. 2 visualizes the distribution of probabilities per dementia syndrome in the form of spider plots. Syndromes with highly specific

Table 3
Model metrics for multi-syndrome classification.

	Sensitivity (Recall)	Positive predictive value (Precision)	F-score
AD	0.42	0.43	0.42
bvFTD	0.60	0.51	0.55
CBS	0.08	0.18	0.11
lvPPA	0.20	0.40	0.27
nfvPPA	0.36	0.34	0.35
PSP	0.54	0.51	0.53
svPPA	0.65	0.61	0.63

Abbreviations: AD Alzheimer's disease; bvFTD behavioral variant frontotemporal dementia; CBS corticobasal syndrome; lvPPA logopenic variant primary progressive aphasia; nfvPPA nonfluent variant primary progressive aphasia; PSP progressive supranuclear palsy; svPPA semantic variant primary progressive aphasia.

Table 2
Binary classification of respective clinical syndrome vs healthy controls.

	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Model accuracy	Balanced Accuracy	Matthew's Correlation Coefficient
AD vs controls	0.89	0.90	0.93	0.85	0.89	0.90	0.79
bvFTD vs controls	0.88	0.63	0.87	0.64	0.81	0.75	0.51
CBS vs controls	0.54	0.80	0.58	0.77	0.71	0.67	0.35
lvPPA vs controls	0.80	0.88	0.80	0.88	0.85	0.84	0.68
nfvPPA vs controls	0.74	0.88	0.83	0.74	0.78	0.78	0.56
PSP vs controls	0.81	0.88	0.87	0.83	0.85	0.85	0.70
svPPA vs controls	0.93	0.96	0.96	0.94	0.95	0.95	0.90

Abbreviations: AD Alzheimer's disease; bvFTD behavioral variant frontotemporal dementia; CBS corticobasal syndrome; lvPPA logopenic variant primary progressive aphasia; nfvPPA nonfluent variant primary progressive aphasia; PSP progressive supranuclear palsy; svPPA semantic variant primary progressive aphasia.

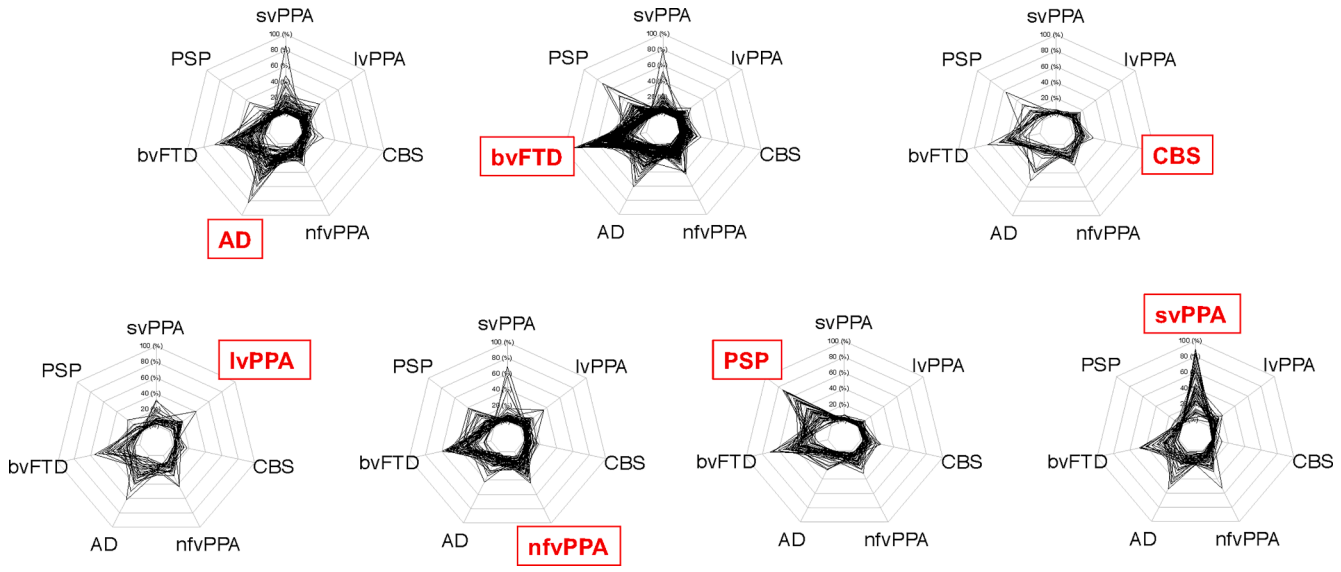

Fig. 2. Spider plots with probability (in %) for different syndromes. Groups of patients are plotted together, and the respective correct diagnosis based on clinical criteria is written in red bold letters and framed within a box. Abbreviations: AD Alzheimer's disease; bvFTD behavioral variant frontotemporal dementia; CBS corticobasal syndrome; lvPPA logopenic variant primary progressive aphasia; nfvPPA nonfluent variant primary progressive aphasia; PSP progressive supranuclear palsy; svPPA semantic variant primary progressive aphasia.

atrophy patterns such as svPPA, bvFTD and PSP show the best distinction from other syndromes. Non-specific atrophy patterns, such as the pattern associated with CBS, show a high variability in probability distribution over different syndromes, potentially reflecting the non-specificity of the pattern. Note the probabilistic gravitation towards the groups with larger sizes as illustrated in Table 4.

Placeholder Fig. 2.

3.6. *Syndrome probabilities in relation to duration and severity of disease*

Probabilities of every patient for the specific clinical syndrome, which the patient was actually suffering from, was related to severity (here FTLD-CDR) and duration of the respective disease (years). This analysis aimed at identifying the impact of disease duration and disease severity on correct syndrome classification based on volumetric MRI. This analysis was performed for both, binary classification, i.e., disease vs controls, and for multiclass prediction, i.e., disease vs disease. Fig. 3 illustrates the results. Remarkably, the analysis revealed associations between probability classified as correct disease and duration and severity of this disease. For binary classification, we hypothesized that one disease can be detected better with stronger disease severity and longer disease duration based on increasing global atrophy. Indeed, generally probability of correct disease classification increased with stronger disease severity and longer disease duration. Fig. 3 shows results of linear regression analysis with significant effects for disease severity in bvFTD and AD ($p < 0.05$; for bvFTD still significant after Bonferroni correction for multiple comparisons), presumably due to high statistical power in these large patient groups.

Another analysis was conducted for multi-syndrome classification. Here, we hypothesized an optimal window with disease-specific clinical symptoms and disease-specific regional brain atrophy, where clinical disease symptoms and related atrophy were strong enough to be detected in comparison to earliest stages, but not that strong involving all cognitive functions and whole-brain atrophy as expected in later disease stages. As we did not expect a linear relationship, we approximated a trend line using local polynomial regression. Results are shown in Fig. 3 again. In agreement with our assumption, best classification was achieved for approximately 2.5 years disease duration in AD, bvFTD, CBS, nfvPPA and svPPA. For lvPPA, we did not observe clear peaks. For PSP, classification probability showed an u-curve pattern, i.e., best classification probability with short disease duration, lowest classification probability with 3 to 5 years, and later increase of probability. For disease severity, as measured with the FTLD-CDR, best classification probabilities were obtained with a score between 2.5 and 5, here for AD, PSP, nfvPPA, svPPA, and lvPPA. CBS was classified best (but still rather incorrect) with an FTLD-CDR score of around 7.5. In bvFTD, classification probability was homogeneously strong with a FTLD-CDR from 0 to 10, and increased further later on. Of note, range of disease duration and disease severity varied remarkably between syndromes, and longer disease duration and higher FTLD-CDR scores were observed in smaller numbers of subjects only. Accordingly, we focus on the range available for almost all syndromes.

Table 4
Mean probabilities in % with standard deviation for every syndrome.

	Probability						
Diagnosis	AD	bvFTD	CBS	lvPPA	nfvPPA	PSP	svPPA
AD	**28 ± 15.6**	31.6 ± 14	5.2 ± 4.6	8.1 ± 5.9	11.5 ± 6.7	5.6 ± 5	10 ± 13.5
bvFTD	14.4 ± 16.1	**44.4 ± 11**	6.22 ± 8	5.4 ± 3.7	13.5 ± 8	9.9 ± 11	6.1 ± 9.5
CBS	14.7 ± 13.6	35.6 ± 11.1	**11 ± 5.1**	7.3 ± 3.5	16 ± 6.5	13.2 ± 11.4	2.2 ± 1.5
lvPPA	23.5 ± 14.2	26.1 ± 13.1	5.8 ± 4	**12.7 ± 7.4**	17.8 ± 9	4.6 ± 5.2	9.6 ± 8
nfvPPA	13.2 ± 8.7	34.6 ± 13	6.7 ± 4	8.8 ± 6.6	**19.5 ± 8.8**	9.1 ± 9.4	8 ± 12.9
PSP	7.9 ± 6.9	36.9 ± 13.9	9 ± 5.5	3.8 ± 2.6	11.8 ± 7.2	**28.8 ± 19.2**	1.9 ± 1.8
svPPA	18.1 ± 12.7	18.3 ± 13.5	2 ± 2.3	8.3 ± 4.7	12.4 ± 9.6	2.8 ± 2.5	**38.2 ± 24.3**

Note: Data are reported as mean ± standard deviation. Abbreviations: AD Alzheimer's disease; bvFTD behavioral variant frontotemporal dementia; CBS corticobasal syndrome; lvPPA logopenic variant primary progressive aphasia; nfvPPA nonfluent variant primary progressive aphasia; PSP progressive supranuclear palsy; svPPA semantic variant primary progressive aphasia.

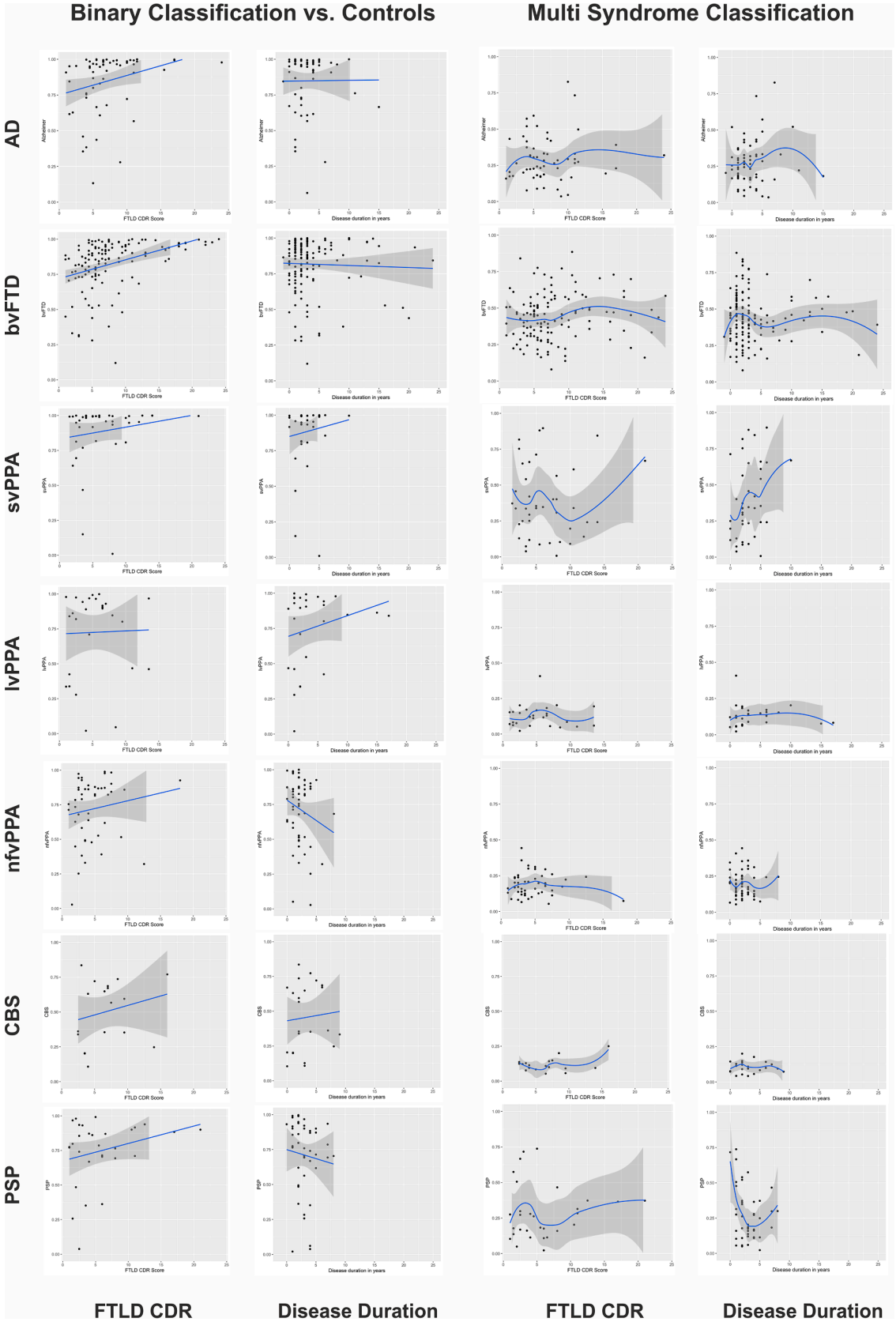

Fig. 3. Syndrome probabilities for correct disease classification in relation to severity and duration of disease. Scatterplots are shown for binary classification, i.e., respective disease vs controls (left), and differential diagnostic multiclass prediction, i.e., disease vs other diseases (right). Linear regression (left) or local polynomial regression (right). Abbreviations: AD Alzheimer's disease; bvFTD behavioral variant frontotemporal dementia; CBS corticobasal syndrome; FTLD-CDR frontotemporal lobar degeneration-modified clinical dementia rating scale; lvPPA logopenic variant primary progressive aphasia; nfvPPA nonfluent variant primary progressive aphasia; PSP progressive supranuclear palsy; svPPA semantic variant primary progressive aphasia.

4. Discussion

To the best of our knowledge this is one of the first studies assessing computerized methods to differentiate multiple (here seven) dementia syndromes based on atrophy patterns with MRI-derived volumetric data of the brain and SVM. One other study applied comparable methods to distinguish several motor syndromes (here three to four typical and atypical Parkinsonian syndromes) (Huppertz et al., 2016), another rather technical recent study from our group compared the performance of different machine learning algorithms in differential diagnosis of neurodegenerative diseases (Lampe et al., 2022). Furthermore, binary classifiers were constructed to differentiate each of the seven dementia syndromes against healthy controls. Our study included multicenter data from 426 well characterized patients suffering from AD, bvFTD, CBS, lvPPA, nfvPPA, PSP, svPPA, and 51 healthy controls. To simulate real-world circumstances, we used the LOOCV approach, where the patient to be predicted was unknown to the classifier, which was trained on the remaining data of the sample. This assured the generalizability of the model performance measures to out-of-sample data.

The binary syndrome models reached high prediction accuracies between 71 and 95 % with a chance level of 50 % as a mathematical comparator. Results suggest that differentiation between dementia syndromes and healthy controls is ready for translation to clinical routine, if validated in external prospective cohorts in the future. The multi-syndrome model reached accuracies of more than three times higher than chance level as a mathematical comparator but was far from 100 %. Accordingly, our results for multi-syndrome classification are promising as a proof-of-principle study but not translatable to clinical settings yet. Multi-syndrome model performance was not homogenous across dementia syndromes. Overall, results suggest that automated methods could support early diagnosis and facilitate appropriate therapy. Furthermore, orphan diseases can be considered by computerized methods, which would avoid indicative imaging biomarkers to be overseen due to a lack of specialized knowledge. In the following we discuss, firstly, results for binary syndrome classifiers, and, secondly, results for multi-syndrome classifiers in more detail.

4.1. Disease-predictive brain regions

The calculation of variable regional importance of a binary SVM classifier that discerned patients with a dementia syndrome from healthy controls coincided with the regions extracted in meta-analyses as disease-specific (Fig. 1). Remarkably, well-known canonical atrophy patterns were reproduced as predictive markers such as lower volume of hippocampal and temporal structures for AD (Henneman et al., 2009), lower volume of the midbrain for PSP (Oba et al., 2005), lower volume in the frontal lobe for bvFTD (Rascovsky et al., 2011, Schroeter et al., 2014), and lower volume of left-sided disease-specific structures for the three PPAs (Rascovsky et al., 2011). Interestingly, a previous meta-analytical study suggested the overlap of left hippocampal structures as a shared regional atrophy between patients with svPPA and AD (Schroeter et al., 2011), which is now reflected in the suggested atrophy patterns in Fig. 1. Results confirm the suggested diagnostic imaging biomarkers for these diseases, i.e., AD, bvFTD and PPAs, where integration into diagnostic criteria is still a desideratum for PSP in the future. On the other hand, CBS seems to be a disease not characterized by focused brain atrophy in agreement with meta-analyses (Albrecht et al., 2017), making the development of diagnostic imaging markers a challenge. More information on the included CBS cohort in our study can be found elsewhere (Albrecht et al., 2019; Ballarini et al., 2020).

4.2. Multi-syndrome prediction

While binary SVM classifications (patients vs controls) rendered very high performance measures consistent with previous studies, multi-syndrome classification is necessary for differential diagnosis. In the daily routine, the radiologist is confronted with more than the question of whether a brain image indicates healthy aging versus atrophy. More importantly, the atrophy pattern should be viewed in the context of different neurodegenerative atrophy patterns that might serve as a morphological indicator for specific dementia syndromes. This is particularly challenging when dealing with rare forms of dementia such as the subtypes of PPA, which have been redefined recently (Gorno-Tempini et al., 2011). While pathognomonic changes such as hippocampal atrophy associated with AD are easily detected, more subtle patterns require a holistic assessment of every brain region to find the best matching diagnosis. The latter is challenging, and computational algorithms might provide valuable diagnostic aid here (Klöppel et al., 2008).

The multi-syndrome classifier used in this study rendered good classification results. Dementia syndromes with more specific atrophy patterns were classified best i.e., svPPA, bvFTD, and (the primary motor syndrome) PSP. This finding is relevant for clinical diagnostics as these syndromes are related to rare orphan diseases that might be missed by physicians in clinical routine. However, not every entity rendered high performance measures. Prediction measures for CBS, a form of neurodegeneration with less region-specific atrophy pattern and divergent underlying histopathology (Albrecht et al., 2017), were weak. Note that the weak classification results of CBS despite a nonspecific atrophy pattern might be also related to the small sample size in our cohort of only 26 patients. On the other hand, it was not the largest patient group available that reached the best classification results. BvFTD and AD comprised the largest groups of patients, yet classification of svPPA and PSP outperformed the classification of bvFTD and AD, indicating that group size might not be the most important factor.

Multi-syndrome prediction not only aids the process of differential diagnosis, it also provides the opportunity to compare probabilities of all included syndromes per individual. We created spider plots to visualize the distribution of syndrome probabilities per groups of patients. While we treated syndromes as categorical concepts during classification, the reality blends different syndromes together. LvPPA for example most often is a subform of AD (Bonner et al., 2010). Remarkably, patients with lvPPA had high probabilities for AD, suggesting a resemblance of atrophy patterns. Moreover, the clinical syndrome CBS might be related histopathologically to AD beside corticobasal degeneration, PSP and other diseases, which might have diminished accuracy in identification of CBS in our study (Franzmeier et al., 2022; Koga et al., 2022). NfvPPA in turn has been associated mainly with tau pathology that is also relevant for bvFTD (Bonner et al., 2010), which is reflected in the probability distribution of patients with the diagnosis nfvPPA. Multi-syndrome prediction with probability distribution allows for detecting cross-syndrome atrophy patterns that might be related to the same underlying pathology. For example, patients with the diagnosis PSP had high probabilities for bvFTD, both syndromes being associated with tauopathies (Meeter et al., 2017). However, note that the probability distribution shows a gravitation towards those patient groups that were most abundant (i.e., bvFTD), a problem that could be overcome with even groups sizes. Histopathological / etiological overlap between clinical syndromes and uncharacteristic atrophy patterns might have led to suboptimal results for the multi-syndrome classifier hampering translation to clinical settings. Furthermore, SVM only allows for post-hoc probability calculation; other machine learning methods should be considered in the future to determine whether they are more suited to calculate accuracy, such as deep neural networks or ensemble methods (Lampe et al., 2022).

Other studies with MRI used combined features quantifying volumetric and morphometric characteristics from T1, and vascular characteristics from fluid-attenuated inversion recovery (FLAIR) images (Koikkalainen et al., 2016). Multi-class classifiers based on disease state index methodology could distinguish between AD, FTD, vascular dementia, dementia with Lewy bodies, and control subjects with a cross-validated classification accuracy of 70.6 % and a balanced accuracy of

69.1 %, better than classification accuracies obtained with visual MRI ratings (accuracy 44.6 %, balanced accuracy 51.6 %). Another study by the same group and based on the same data (Tong et al., 2017) confirmed this finding by using random undersampling boosting. Here, additionally a group with subjective memory complaints and measures from cerebrospinal fluid had been included. The authors achieved a high accuracy of 75.2 % and balanced accuracy of 69.3 %. These results underline the translatability of machine learning on MRI to clinical settings and the importance of multimodal approaches, tailored to clinical needs.

Moreover, we investigated – for both, binary classification vs controls, and multiclass prediction across diseases – optimal time and severity windows for disease identification with structural MRI. Here, we related classification probabilities of every patient for the specific clinical syndrome the patient actually was suffering from, to severity and duration of the respective disease. This analysis aimed at identifying the impact of disease duration and disease severity on correct syndrome classification based on volumetric MRI. For binary classification, we confirmed our hypothesis that one disease can be detected better with stronger disease severity and longer disease duration based on increasing global atrophy.

For multiclass classification, i.e., disease vs disease, we detected for most of the syndromes an optimal diagnostic window with already disease-specific clinical symptoms, and, presumably disease-specific regional brain atrophy, where clinical disease symptoms and related atrophy were strong enough to be detected in comparison to earliest stages, but not that severe such as in later disease stages, where all cognitive functions are involved and atrophy already has spread across the whole brain. The optimal diagnostic window for classification based on MRI was approximately 2.5 years for disease duration. In contrast, PSP showed best classification probability with short disease duration, lowest classification probability with 3 to 5 years, and later again an increase of probability. For disease severity, best classification probabilities were obtained mainly with a score between 2.5 and 5, for CBS 7.5, whereas for bvFTD classification probability was homogeneously strong. In sum, this analysis revealed specific time and severity windows for specific diseases, where the MRI-based classifier can best detect the specific disease, which is of particular importance if the classifier if applied in clinical settings.

4.3. Limitations & future perspectives

A limitation of this dataset is the uneven distribution of the different groups. Because we wanted to include the maximum amount of data, we used every available patient, which resulted in an unbalanced dataset (e.g., 146 patients with bvFTD contrasting only 26 patients with CBS). Distribution of subjects included in each diagnostic group mirrors the clinical incidence / prevalence of the different neurodegenerative syndromes. Undersampling the larger groups bore the risk of leaving out important information to build the classifier, while oversampling of the smaller groups would have led to overfitting the model. Nonetheless, imbalanced training datasets might lead to classifiers biased towards the majority class. To account for these limitations, we calculated model performance measures considering this imbalance such as the Matthews correlation coefficient for binary classification. Ideally, there will be more evenly distributed and especially larger datasets in the future, enabling even validation in independent cohorts.

Moreover, one might discuss a potential age bias in our results. Although disease cohorts differed in mean age, the several disease cohorts were age-matched with healthy controls, i.e., both respective groups did not differ in age significantly. The binary classification also identified brain regions as classification-relevant that are well known as disease-specific from the literature. Accordingly, we consider an age bias as not relevant in our study, at least for the binary syndrome classification. Although mean age differences between groups potentially might have biased results for the multiple syndrome classification, mean age differences were small (ranging from 62 to 69 years) and mean disease duration was balanced between groups. Beside age, one might discuss sex as a potential bias. Of note, a recent study has investigated the impact of the covariates total intracranial volume, age and sex on voxel based morphometry results in another neurodegenerative disease, Parkinson's disease (Crowley et al., 2018). Remarkably, total intracranial volume had an impact on results, whereas the covariates age and sex were negligible. Remarkably, normalizing for total intracranial volume has been reported to abolish the majority of sex-related volume differences (Kijonka et al., 2020; Sanchis-Segura et al., 2020). Although, based on these studies, we regard controlling for total intracranial volume as most important, future studies shall investigate the potential impact of age and sex / gender more cautiously. Differences in disease severity measures such as MMSE, CDR and FTLD-CDR might be related to the well-known fact that a measure might be more sensitive for one specific disease, such as the CDR and MMSE for AD. Developing disease-spanning tools for disease severity is a desideratum for the future, which will be hampered by disparate symptom-profiles.

One might also criticize the relatively small size of some patient cohorts and the control sample. Particularly low group numbers of lvPPA, CBS and PSP might be related to differences in prevalence / incidence between diseases as discussed above. As healthy controls had to be investigated with the same scanners / parameters to guarantee comparability we could not extend this cohort by including external data. Note that the controls' sample size is only relevant for the binary classifiers, which yielded excellent accuracies. In the multi-syndrome classifier controls were not included. To adjust for this possible inhomogeneity bias, we also used statistical approaches to account for uneven sample sizes (Matthews Correlation Coefficient). Future data sets shall contain larger numbers of controls as low numbers of healthy subjects might generally penalize results.

Although different MRI machines and parameters for data acquisition per center might yield another bias, we assume that the method applied, i.e. atlas based volumetry, controlled for this difference, at least partly, by normalizing the volumes of the atlas structures to total intracranial volume. However, this assumption has to be validated in another study. Furthermore, elderly patients have age-related comorbidities that might affect brain structure independent of the underlying neurodegenerative disease adding to the complexity of the data. While patients with major brain pathologies such as stroke or brain tumors were excluded, vascular changes such as white matter hyperintensities possibly affecting both cognition and brain structure might be a confounding factor as well (Lampe et al., 2019).

Machine learning classification depends on the quality of data to detect relevant patterns. We treated dementia syndromes as singular independent entities that exclude each other. Even though this is conceptually possible in an ideal clinical world, in the real world this is not the case. In fact, clinical dementia syndromes with the same manifestation can be related to different underlying pathologies (Meeter et al., 2017). While this complex relationship can be acknowledged with the syndrome probabilities, it necessarily corrupts the performance measures of the classification itself. These factors shall be controlled for in future studies. As our classifier's performance might be overestimated with the current approach it has to be proven in an independent validation cohort. Furthermore, a successful transition of the classifier to prospective clinical data has yet to be tested. The model remains to be validated on data from different field strengths and lower quality, in cases, where diagnosis had been histopathologically validated, and independent datasets including also other neurodegenerative diseases such as Lewy body dementia/disease and posterior cortical atrophy. Besides building classifiers for differentiation between multiple dementia / neurodegenerative syndromes, as done in our study, specifically tailored classifiers for differential diagnosis might be designed adapted to clinical needs by including clinically related syndromes only, such as differential diagnosis of the three PPA subtypes (Bisenius et al., 2017) or of typical and atypical Parkinsonian syndromes (Huppertz

et al., 2016).

Conceptually, we were interested in using MRI data for disease classification only. Future studies might evaluate the additional benefit of imaging data if parameters from neuropsychological / patholinguistic tests and behavioral questionnaires investigating, for instance apathy or behavioral changes, are used for disease classification. Then, classification accuracies of multiple syndrome models shall be compared with other comparators also beside mathematical ones as done in our study, such as classification accuracies by clinicians including all available clinical / biomarker information or radiologists' reads. Finally, multimodal imaging data might much better reflect the multimodal nature of neurodegeneration and might lead to better classification results (Pievani et al., 2011). Therefore, other imaging modalities, such as connectivity measures or molecular imaging, shall be included beside structural MRI, which might yield higher classification accuracies (Dukart et al., 2011). Importantly, histopathology underlying neurodegenerative diseases and systematic genetic analyses shall be taken into account in following studies. These measures might include surrogate markers from cerebrospinal fluid or positron emission tomography, such as amyloid, tau, phospho tau, and neurofilaments, or genetics and post mortem validation. Herewith, accuracy might be increased beyond simple atrophy patterns, where obviously some dementia types might show overlapping patterns.

5. Author credits

LL and MLS conceptualized the study. LL wrote the first draft of the paper together with MLS. LL and HJH analyzed data, supervised by MLS and KM. SAS, FA, TB, SB, KM, KFa, KFl, HJ, JH, ML, JP, AS, MS, JK, AD, JDS, MO were involved in data collection. MLS, LL, HJH and SN revised the paper. AV supported the study regarding funding. All authors approved the final and revised version of the paper.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This work was supported by the German Federal Ministry of Education, and Research (BMBF) by a grant given to the German FTLD Consortium (FKZ O1GI1007A), by the German Research Foundation (DFG, SCHR 774/5-1), by the Parkinson's Disease Foundation (PDF-IRG-1307), by the Michael J. Fox Foundation (MJFF-11362), and the eHealthSax Initiative of the Sächsische Aufbaubank (SAB; project TelDem). Accordingly, this study is co-financed with tax revenue based on the budget approved by the Saxon state parliament.

References

Albrecht, F., Bisenius, S., Morales Schaack, R., Neumann, J., Schroeter, M.L., 2017. Disentangling the neural correlates of corticobasal syndrome and corticobasal degeneration with systematic and quantitative ALE meta-analyses. NPJ Parkinsons Dis 3, 12.

Albrecht, F., Mueller, K., Ballarini, T., Lampe, L., Diehl-Schmid, J., Fassbender, K., et al., 2019. Unraveling corticobasal syndrome and alien limb syndrome with structural brain imaging. Cortex 117, 33–40.

Anderl-Straub S, Lausser L, Lombardi J, Uttner I, Fassbender K, Fliessbach K, H.J. Huppertz et al. Predicting disease progression in behavioral variant frontotemporal dementia Alzheimers Dement (Amst). 13 1 2021 10.1002/dad2.12262 e12262.

Arbabshirani, M.R., Plis, S., Sui, J., Calhoun, V.D., 2017. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. Neuroimage 145, 137–165.

Ballarini, T., Albrecht, F., Mueller, K., Jech, R., Diehl-Schmid, J., Fliessbach, K., et al., 2020. Disentangling brain functional network remodeling in corticobasal syndrome – A multimodal MRI study. Clinical, NeuroImage.

Bisenius, S., Neumann, J., Schroeter, M.L., 2016. Validating new diagnostic imaging criteria for primary progressive aphasia via anatomical likelihood estimation meta-analyses. Eur. J. Neurol. 23, 704–712.

Bisenius, S., Mueller, K., Diehl-Schmid, J., Fassbender, K., Grimmer, T., Jessen, F., et al., 2017. Predicting primary progressive aphasias with support vector machine approaches in structural MRI data. Neuroimage Clin 14, 334–343.

Bonner, M.F., Ash, S., Grossman, M., 2010. The New Classification of primary progressive aphasia into semantic, logopenic, or nonfluent/agrammatic variants. Curr. Neurol. Neurosci. Rep. 10, 484–490.

Brant-Zawadzki, M., Gillan, G.D., Nitz, W.R., 1992. MP RAGE: A three-dimensional, T1-weighted, gradient-echo sequence - initial experience in the brain. Radiology 182, 769–775.

Brodersen KH, Ong CS, Stephan KE, Buhmann JM. The balanced accuracy and its posterior distribution. Proceedings of the 20th International Conference on Pattern Recognition 2010, 3121-24.

Bron, E.E., Smits, M., Niessen, W.J., Klein, S., 2015. Feature Selection Based on the SVM Weight Vector for Classification of Dementia. IEEE J. Biomed. Health Inform. 19 (5), 1617–1626.

Crowley, S., Huang, H., Tanner, J., et al., 2018. Considering total intracranial volume and other nuisance variables in brain voxel based morphometry in idiopathic PD. Brain Imaging Behav. 12, 1–12. https://doi.org/10.1007/s11682-016-9656-9.

De Fauw, J., Ledsam, J.R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., et al., 2018. Clinically applicable deep learning for diagnosis and referral in retinal disease. Nat. Med. 24 (9), 1342–1350. https://doi.org/10.1038/s41591-018-0107-6.

Dukart, J., Mueller, K., Horstmann, A., Barthel, H., Möller, H.E., Villringer, A., et al., 2011. Combined Evaluation of FDG-PET and MRI improves detection and differentiation of dementia. PLoS One 6, e18111.

Dwyer, D.B., Falkai, P., Koutsouleris, N., 2018. Machine learning approaches for clinical psychology and psychiatry. Annu. Rev. Clin. Psychol. 14, 91–118.

Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., et al., 2017. Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115–118.

Franzmeier, N., Koutsouleris, N., Benzinger, T., Goate, A., Karch, C.M., Fagan, A.M., et al., 2020. Predicting sporadic Alzheimer's disease progression via inherited Alzheimer's disease-informed machine-learning. Alzheimers Dement. 16 (3), 501–511. https://doi.org/10.1002/alz.12032.

Franzmeier, N., Brendel, M., Beyer, L., et al., 2022. Tau deposition patterns are associated with functional connectivity in primary tauopathies. Nat. Commun. 13, 1362.

Goins, R.T., Williams, K.A., Carter, M.W., Spencer, M., Solovieva, T., 2005. Perceived barriers to health care access among rural older adults: a qualitative study. J. Rural Health 21, 206–213.

Gorno-Tempini, M.L., Hillis, A.E., Weintraub, S., Kertesz, A., Mendez, M., Cappa, S.F., et al., 2011. Classification of primary progressive aphasia and its variants. Neurology 76, 1006–1014.

Henneman, W.J.P., Sluimer, J.D., Barnes, J., van der Flier, W.M., Sluimer, I.C., Fox, N.C., et al., 2009. Hippocampal atrophy rates in Alzheimer disease: Added value over whole brain volume measures. Neurology 72, 999–1007.

Huppertz, H.-J., Möller, L., Südmeyer, M., Hilker, R., Hattingen, E., Egger, K., et al., 2016. Differentiation of neurodegenerative parkinsonian syndromes by volumetric magnetic resonance imaging analysis and support vector machine classification. Mov. Disord. 31, 1506–1517.

Kijonka, M., Borys, D., Psiuk-Maksymowicz, K., Gorczewski, K., Wojcieszek, P., Kossowski, B., et al., 2020. Whole brain and cranial size adjustments in volumetric brain analyses of sex- and age-related trends. Front. Neurosci. 14, 278. https://doi.org/10.3389/fnins.2020.00278.

Klöppel, S., Stonnington, C.M., Barnes, J., Chen, F., Chu, C., Good, C.D., et al., 2008. Accuracy of dementia diagnosis - A direct comparison between radiologists and a computerized method. Brain 131, 2969–2974.

Koga S, Josephs KA, Aiba I, Yoshida M, Dickson DW. Neuropathology and emerging biomarkers in corticobasal syndrome J Neurol Neurosurg Psychiatry 93 9 2022 Jun 13 919 929 10.1136/jnnp-2021-328586.

Koikkalainen, J., Rhodius-Meester, H., Tolonen, A., Barkhof, F., Tijms, B., Lemstra, A.W., et al., 2016. Differential diagnosis of neurodegenerative diseases using structural MRI data. Neuroimage Clin 5 (11), 435–449.

Koutsouleris, N., Pantelis, C., Velakoulis, D., McGuire, P., Dwyer, D.B., Urquijo-Castro, M.F., et al., 2022. Exploring links between psychosis and frontotemporal dementia using multimodal machine learning: Dementia praecox revisited. JAMA Psychiat. 79 (9), 907–919. https://doi.org/10.1001/jamapsychiatry.2022.2075.

Lampe, L., Kharabian-Masouleh, S., Kynast, J., Arelin, K., Steele, C.J., Löffler, M., et al., 2019. Lesion location matters: The relationships between white matter hyperintensities on cognition in the healthy elderly. J. Cereb. Blood Flow Metab. 39 (1), 36–43. https://doi.org/10.1177/0271678X17740501.

Lampe, L., Niehaus, S., Huppertz, H.J., et al., 2022. Comparative analysis of machine learning algorithms for multi-syndrome classification of neurodegenerative syndromes. Alz Res Therapy 14, 62.

Lin, H.-T., Lin, C.-J., Weng, R.C., 2007. A note on Platt's probabilistic outputs for support vector machines. Mach. Learn. 68, 267–276.

Magnin, B., Mesrob, L., Kinkingnéhun, S., Pélégrini-Issac, M., Colliot, O., Sarazin, M., et al., 2009. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology 51, 73–83.

Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. BBA 405, 442–451.

McKhann, G.M., Knopman, D.S., Chertkow, H., Hyman, B.T., Jack, C.R., Kawas, C.H., et al., 2011. The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. 7, 263–269.

Meeter, L.H., Kaat, L.D., Rohrer, J.D., van Swieten, J.C., 2017. Imaging and fluid biomarkers in frontotemporal dementia. Nat. Rev. Neurol. 13, 406–419. https://doi.org/10.1038/nrneurol.2017.75.

Meyer, S., Mueller, K., Stuke, K., Bisenius, S., Diehl-Schmid, J., Jessen, F., et al., 2017. Predicting behavioral variant frontotemporal dementia with pattern classification in multi-center structural MRI data. Neuroimage Clin 14, 656–662.

Mueller, K., Jech, R., Bonnet, C., Tintěra, J., Hanuška, J., Möller, H.E., et al., 2017. Disease-specific regions outperform whole-brain approaches in identifying progressive supranuclear palsy: A multicentric MRI study. Front. Neurosci. 11, 100.

Oba, H., Yagishita, A., Terada, H., Barkovich, A.J., Kutomi, K., Yamauchi, T., et al., 2005. New and reliable MRI diagnosis for progressive supranuclear palsy. Neurology 64, 2050–2055.

Otto, M., Ludolph, A.C., Landwehrmeyer, B., Förstl, H., Diehl-Schmid, J., Neumann, M., et al., 2011. German consortium for frontotemporal lobar degeneration. Nervenarzt 82, 1002–1005.

Pievani, M., de Haan, W., Wu, T., Seeley, W.W., Frisoni, G.B., 2011. Functional network disruption in the degenerative dementias. Lancet Neurol. 10 (9), 829–843.

Rascovsky, K., Hodges, J.R., Knopman, D., Mendez, M.F., Kramer, J.H., Neuhaus, J., et al., 2011. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain 134, 2456–2477.

Sanchis-Segura, C., Ibañez-Gual, M.V., Aguirre, N., et al., 2020. Effects of different intracranial volume correction methods on univariate sex differences in grey matter volume and multivariate sex prediction. Sci. Rep. 10, 12953.

Schölkopf, B., Burges, C.J.C., Smola, A.J. (Eds.), 1999. Advances in Kernel Methods: Support Vector Learning. MIT Press, Cambridge, MA, USA.

Schroeter, M.L., Laird, A.R., Chwiesko, C., Deuschl, C., Schneider, E., Bzdok, D., et al., 2014. Conceptualizing neuropsychiatric diseases with multimodal data-driven meta-analyses - The case of behavioral variant frontotemporal dementia. Cortex 57, 22–37.

Schroeter, M.L., Neumann, J., 2011. Combined Imaging Markers Dissociate Alzheimer's Disease and Frontotemporal Lobar Degeneration - An ALE Meta-Analysis. Front. Aging Neurosci. 3, 10.

Schroeter, M.L., Raczka, K., Neumann, J., Yves von Cramon, D., 2007. Towards a nosology for frontotemporal lobar degenerations - A meta-analysis involving 267 subjects. Neuroimage 36, 497–510.

Schroeter, M.L., Raczka, K., Neumann, J., von Cramon, D.Y., 2008. Neural networks in frontotemporal dementia - A meta-analysis. Neurobiol. Aging 29, 418–426.

Schroeter, M.L., Stein, T., Maslowski, N., Neumann, J., 2009. Neural correlates of Alzheimer's disease and mild cognitive impairment: A systematic and quantitative meta-analysis involving 1351 patients. Neuroimage 47, 1196–1206.

Shattuck, D.W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K.L., et al., 2008. Construction of a 3D probabilistic atlas of human cortical structures. Neuroimage 39, 1064–1080.

Tahmasian, M., Shao, J., Meng, C., Grimmer, T., Diehl-Schmid, J., Yousefi, B.H., et al., 2016. Based on the Network degeneration hypothesis: Separating individual patients with different neurodegenerative syndromes in a preliminary hybrid PET/MR study. J. Nucl. Med. 57 (3), 410–415.

Tong, T., Ledig, C., Guerrero, R., Schuh, A., Koikkalainen, J., Tolonen, A., et al., 2017. Five-class differential diagnostics of neurodegenerative diseases using random undersampling boosting. Neuroimage Clin 15, 613–624.

Vapnik VN. The nature of statistical learning theory. 1995.; https://cran.r-project.org/web/packages/e1071/e1071.pdf; https://www.csie.ntu.edu.tw/~cjlin/libsvm/.

Wagner, M., Lorenz, G., Volk, A.E., et al., 2021. Clinico-genetic findings in 509 frontotemporal dementia patients. Mol. Psychiatry 26, 5824–5832.

Warren, J.D., Fletcher, P.D., Golden, H.L., 2012. The paradox of syndromic diversity in Alzheimer disease. Nat. Rev. Neurol. 8, 451–464. https://doi.org/10.1038/nrneurol.2012.135.

Whitwell, J.L., Höglinger, G.U., Antonini, A., Bordelon, Y., Boxer, A.L., Colosimo, C., et al., 2017. Radiological biomarkers for diagnosis in PSP: Where are we and where do we need to be? Mov. Disord. 32, 955–971.

Woo, C.-W., Chang, L.J., Lindquist, M.A., Wager, T.D., 2017. Building better biomarkers: Brain models in translational neuroimaging. Nat. Neurosci. 20, 365–377.

Yu, K.-H., Zhang, C., Berry, G.J., Altman, R.B., Ré, C., Rubin, D.L., et al., 2016. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. Nat. Commun. 7, 12474.

Zheng, C., Xia, Y., Pan, Y., Chen, J., 2016. Automated identification of dementia using medical imaging: A survey from a pattern classification perspective. Brain Inform 3 (1), 17–27.