

# A multimodal neuroimaging classifier for alcohol dependence

Matthias Guggenmos<sup>1</sup>, Katharina Schmack<sup>1</sup>, Ilya M. Veer<sup>1</sup>, Tristram Lett<sup>1</sup>, Maria Sekutowicz<sup>1</sup>, Miriam Sebold<sup>1</sup>, Maria Garbusow<sup>1</sup>, Christian Sommer<sup>2</sup>, Hans-Ulrich Wittchen<sup>4,5</sup>, Ulrich S. Zimmermann<sup>2</sup>, Michael N Smolka<sup>2,3</sup>, Henrik Walter<sup>1</sup>, Andreas Heinz<sup>1</sup>, Philipp Sterzer<sup>1</sup>

<sup>1</sup> Department of Psychiatry and Psychotherapy, Charité Universitätsmedizin, Germany

<sup>2</sup> Department of Psychiatry and Psychotherapy, Technische Universität Dresden, Germany

<sup>3</sup> Neuroimaging Center, Technische Universität Dresden, Dresden, Germany

<sup>4</sup> Institute of Clinical Psychology and Psychotherapy, Technische Universität Dresden, Germany

<sup>5</sup> Department of Psychiatry and Psychotherapy, Ludwig Maximilians Universität Munich, Munich, Germany

## Corresponding author:

Matthias Guggenmos  
Charité Universitätsmedizin Berlin  
Department of Psychiatry and Psychotherapy  
Address: Charitéplatz 1, 10117 Berlin, Germany  
Phone: +49 30 450 517131  
E-Mail: matthias.guggenmos@charite.de

## **Acknowledgments**

This work was supported by the following institutions: German Research Foundation (Deutsche Forschungsgemeinschaft, DFG, FOR 1617: grants STE 1430/6-1, STE 1430/6-2, SCHM 3209/1-2, ZI 1119/3-1, ZI 1119/3-2, HE 2597/14-1, HE 2597/14-2, WI 709/10-1, WI 709/10-2, GU 1845/1-1 and Excellence Cluster Exc 257); Federal Ministry of Education and Research (BMBF grants 01ZX1311H; 01ZX1311D/1611D and 01ZX1311E/1611E; and in part by 01EE1406A and 01EE1406B). K.S. is supported by a research fellowship from the Leopoldina - Research Fellowship of Leopoldina–German National Academy of Sciences (LPDS 2018-03).

## **Author contribution statement**

M. Gu. drafted the manuscript. K. S., M. Sek. and P. S. provided critical revisions of the manuscript. All authors contributed valuable feedback to the manuscript. All authors contributed conceptually to the design of the study and to ideas for analyses.

## **Additional information**

The authors declare no competing interests.

1    **Abstract**

2    With progress in magnetic resonance imaging technology and a broader dissemination of  
3    state-of-the-art imaging facilities, the acquisition of multiple neuroimaging modalities is  
4    becoming increasingly feasible. One particular hope associated with multimodal  
5    neuroimaging is the development of reliable data-driven diagnostic classifiers for psychiatric  
6    disorders, yet previous studies have often failed to find a benefit of combining multiple  
7    modalities. As a psychiatric disorder with established neurobiological effects at several levels  
8    of description, alcohol dependence is particularly well-suited for multimodal classification.  
9    To this aim, we developed a multimodal classification scheme and applied it to a rich  
10   neuroimaging battery (structural, functional task-based and functional resting-state data)  
11   collected in a sample of alcohol-dependent patients (N=119) and controls (N=97). We found  
12   that our classification scheme yielded 79.3% diagnostic accuracy, which outperformed the  
13   strongest individual modality – grey-matter density – by 2.7%. We found that this moderate  
14   benefit of multimodal classification depended on a number of critical design choices: a  
15   procedure to select optimal modality-specific classifiers, a fine-grained ensemble prediction  
16   based on cross-modal weight matrices and continuous classifier decision values. We conclude  
17   that the combination of multiple neuroimaging modalities is able to improve the accuracy of  
18   machine-learning-based diagnostic classification in alcohol dependence, but currently only to  
19   a moderate degree.

20

21 **INTRODUCTION**

22 In recent years, technical advancements in magnetic resonance imaging (MRI) technology  
23 and increasing access to these state-of-the-art MRI facilities for both clinicians and  
24 researchers have nourished the quest for MRI-based diagnostic classifiers of psychiatric  
25 disorders that proceed in an automated and objective manner. In addition, multiple MRI  
26 modalities, including high-resolution structural images, resting-state connectivity maps, white  
27 matter tractography based on diffusion tensor imaging, functional MRI, are now readily  
28 available as part of standard experimental protocols. The hope associated with this approach  
29 referred to as *multimodal MRI* is that measurements targeting different levels of brain  
30 structure and function will, in combination, lead to a breakthrough in the quantitative  
31 characterization of psychiatric disorders<sup>1–3</sup>.

32 A limiting factor in this endeavour is that most psychiatric disorders have an upper bound for  
33 the accuracy of machine-learning-based diagnostic classification imposed by (1) ‘label noise’  
34 of psychiatric diagnoses, evidenced by often poor inter-rater reliabilities<sup>4,5</sup>, and (2) an  
35 intrinsic heterogeneity of psychiatric diagnostic labels themselves<sup>6–8</sup>. It is therefore no  
36 surprise that the most successful applications of (multimodal) MRI-based machine learning  
37 have been achieved in disorders involving clear neurodegenerative effects such as  
38 Alzheimer’s disease<sup>9–12</sup> or multiple sclerosis<sup>13,14</sup>.

39 Here, we reasoned that alcohol dependence is a well-suited psychiatric disorder for  
40 automated diagnostic classification based on multimodal MRI and an ideal test case and  
41 benchmark for methodological developments. This is first and foremost because  
42 neurobiological correlates of alcohol dependence have been established at several levels of  
43 description, including grey-matter loss<sup>15–18</sup>, increased ventricular size / cerebrospinal fluid  
44 concentration<sup>19–21</sup> and aberrant neural reward responses<sup>22–25</sup>. The diversity of effects in  
45 combination with a high degree of replicability make the combination of different MRI  
46 modalities appear particularly powerful in the case of alcohol dependence. In addition,  
47 alcohol dependence, relative to other psychiatric disorders, is a reliable diagnosis<sup>26–28</sup> and  
48 thus a paradigmatic case to gauge the true predictive potential of an MRI-based classifier for  
49 psychiatric diagnosis.

50 Our investigation was based on a sample of 119 alcohol-dependent patients and 97 controls  
51 who underwent an extensive neuroimaging battery including structural<sup>16,29</sup>, functional<sup>30–32</sup>  
52 and resting-state<sup>33</sup> MRI as part of the LeAD study ([www.lead-studie.de](http://www.lead-studie.de); clinical trial number:

53 NCT01679145). Our guiding rationale for the construction of a multimodal diagnostic  
54 classifier was clinical utility and practicability, and thus to leverage on neuroimaging  
55 modalities that are effective and simple to acquire (i.e., relatively short acquisition times and  
56 standard scanning sequences) and to analyse (e.g. no computational cognitive modelling or  
57 other approaches rich on assumptions). Selected modalities comprised grey-matter density,  
58 cerebral spinal fluid and cortical thickness based on structural MRI as well as basic reward  
59 responses and resting-state connectivity based on functional MRI.

60 Key methodological challenges of combining multiple MRI modalities are the heterogeneity  
61 of different modalities in terms of statistical properties or number of predictors, missing data  
62 in one or more modalities, and the relative weighting of each modality. In the present  
63 investigation we thus developed a novel classification framework that (1) combined  
64 modality-specific predictions in an ensemble vote, (2) allowed for emphasizing or de-  
65 emphasizing individual modalities through weighting, (3) considered fine-grained  
66 information from modality-specific classifiers instead of binary labels, and (4) was robust to  
67 missing data in individual modalities. We hypothesized that the combination of MRI  
68 modalities for diagnostic classification of alcohol dependence would outperform any  
69 individual unimodal classification approach.

70 **METHOD**

71 **Participants**

72 This study was conducted as part of the Learning and Alcohol Dependence (LeAD) study, a  
73 German (Berlin, Dresden) program investigating the neurobiological basis of alcohol  
74 dependence ([www.lead-studie.de](http://www.lead-studie.de); clinical trial number: NCT01679145<sup>30,34,35</sup>). We assessed  
75 119 individuals aged 20-65 (18 female) meeting criteria of alcohol dependence according to  
76 ICD-10 and DSM-IV-TR (American Psychiatric Association 2000) and 97 healthy controls  
77 aged 21-65 (16 female) matched in terms of age, gender and smoking (see Table 1).

78 We used the computer-assisted interview version Composite International Diagnostic  
79 Interview (CAPI-CIDI<sup>36,37</sup>) to verify diagnosis criteria of AD in the patient group and to  
80 exclude the possibility of AD in control subjects. For inclusion, individuals with AD had to  
81 meet criteria for AD for at least three years and had to undergo an inpatient detoxification  
82 phase (average duration ± SEM: 22.8 ± 1 days). Exclusion criteria for all subjects were left-  
83 handedness (Edinburgh handedness index below 50<sup>38</sup>), contraindications for MRI, and a  
84 history of or current neurological (including Korsakoff syndrome) or mental disorders  
85 (excluding nicotine dependence in both groups and alcohol abuse in individuals with AD, but  
86 including abuse of other drugs). Mental disorders were assessed according to DSM-IV axis  
87 one as verified by the CAPI-CIDI. It was ensured that all subjects were free of psychotropic  
88 medication (including detoxification treatment) known to interact with the central nervous  
89 system for at least four half-lives. Current non-alcohol drug abuse was confirmed by means  
90 of a dedicated urine test.

91 Note that several sampling characteristics were influenced by the fact that predictors of  
92 relapse, while not part of the present work, were another research focus within the LeAD  
93 study<sup>30</sup>: (1) to have sufficient power for the comparison between relapsers and abstainers, the  
94 AD group was oversampled relative to the HC group; (2) to be able to study relapse  
95 behaviour, detoxification at baseline was an inclusion criterion; (3) to prevent the possibility  
96 that handedness would be unequally distributed across relapsers and abstainers, only right  
97 handers were included. Finally note that matching for smoking implied a higher proportion of  
98 smokers in the HC sample than expected from a representative sample (67% versus a  
99 representative prevalence of 28% in Germany<sup>39</sup>).

100 Participants gave written informed consent. Ethical approval for the study was obtained from  
101 both sites (Ethics committee of the Universitätsklinikum Dresden/Technische Universität  
102 Dresden, EK 228072012; Ethics committee of Charité–Universitätsmedizin Berlin, EA  
103 1/157/11), and procedures were in accordance with the Declaration of Helsinki.

104 **Overview of neuroimaging modalities**

105 To maximize the clinical utility of a multimodal diagnostic classifier, we constructed a  
106 neuroimaging battery of five modalities that was 1) simple to acquire and to analyse, and 2)  
107 based on established effects either in previous literature or on own works within the LeAD  
108 program. From structural MRI measurements, we derived three modalities: (1) grey-matter  
109 density, as damaging effects to grey-matter integrity in chronic alcoholics have been  
110 replicated numerous times in the literature<sup>15,17,18</sup> including our own work in which we also  
111 demonstrated high diagnostic discriminability<sup>16</sup>; (2) cerebrospinal fluid, as increased  
112 ventricular size (or increased cerebrospinal fluid, respectively) is arguably one the most  
113 salient characteristics of alcoholic brains<sup>19–21</sup>; and (3) cortical thickness, an additional specific  
114 marker of grey-matter integrity obtained through surface-based analysis, which has been  
115 successfully used in more recent studies to characterize structural damage in alcohol  
116 dependence<sup>40–42</sup>.

117 From task-based functional MRI we obtained (4) functional activation patterns representing a  
118 basic reward response (outcome versus no outcome), motivated by the fact that aberrant  
119 functional reward responses have been consistently found for alcohol dependence<sup>22–25</sup>.  
120 Finally, from resting-state functional MRI, we derived (5) nucleus accumbens whole-brain  
121 connectivity maps, as one of the key research goals of the LeAD program was testing a  
122 hypothesis about disturbed striato-frontal connectivity underlying the development of  
123 problematic drinking behaviours (this hypothesis has recently been confirmed by authors of  
124 this article<sup>33</sup>). Note that all data were linearly corrected for variance of no interest related to  
125 demographic variables age, gender and site.

126 **MRI data acquisition**

127 Magnetic resonance imaging (MRI) was performed on a 3-Tesla Siemens Trio (Erlangen,  
128 Germany) scanner with a 12-channel head-coil to obtain (1) structural, (2) functional task-  
129 based and (3) functional resting-state MRI data. Structural T<sub>1</sub>-weighted MRI scans were  
130 acquired using a magnetization-prepared rapid gradient echo sequence (repetition time: 1900

131 ms; echo time: 5.25 ms; flip angle: 9°; field of view: 256×256 mm<sup>2</sup>; voxel size: 1mm  
132 isotropic; 192 sagittal slices). Functional (task-based or resting-state) T<sub>2</sub>\*-weighted MRI  
133 scans were acquired using a gradient echo planar imaging sequence (repetition time: 2410  
134 ms; echo time: 25 ms; flip angle: 80°; field of view: 192x192 mm<sup>2</sup>; voxel size: 3x3x2 mm<sup>3</sup>)  
135 comprising 42 slices approximately -25° to the bicommissural plane. Volume-to-volume  
136 movement of more than 3mm translation and/or 2 degrees rotation led to exclusion (9 HC and  
137 12 AD subjects were excluded due to these criteria). All imaging data were screened for  
138 corrupted data or serious acquisition artefacts.

139 **MRI preprocessing and feature preparation**

140 Neuroimaging features from overall five modalities were computed: grey-matter density  
141 (GMD), cerebrospinal fluid (CSF) and cortical thickness (CTH) from structural MRI; a basic  
142 reward response (RWR) signal from task-based functional MRI and nucleus accumbens  
143 connectivity (NAC) from resting-state functional MRI. In the following we describe the steps  
144 involved in preprocessing and feature preparation for each modality.

145 **Structural MRI.** Structural MRI images were processed in two separate analyses streams for  
146 GMD/CSF and CTH. For GMD/CSF, SPM12 (<http://www.fil.ion.ucl.ac.uk/spm>) was used in  
147 combination with voxel-based morphometry (VBM8; <http://dbm.neuro.uni-jena.de/vbm>).  
148 T1-weighted Images were spatially normalized to a Montreal Neurological Institute (MNI)  
149 template and segmented based on tissue types. Unmodulated images representing GMD and  
150 CSF density were smoothed with an 8 mm isotropic Gaussian kernel. To reduce  
151 dimensionality, 110 regional averages for both GMD and CSF were computed based on a  
152 combined cortical and subcortical anatomical brain atlas<sup>16,43</sup>.

153 To obtain estimates of CTH, cortical reconstruction was performed on T1-weighted images  
154 using the FreeSurfer morphometric analysis suite (<http://surfer.nmr.mgh.harvard.edu/>). The  
155 technical details of these procedures are described in prior publications<sup>44,45,54,55,46–53</sup>. CTH  
156 estimates were obtained by calculating the closest distance from the grey/white boundary to  
157 the grey/CSF boundary at each vertex on the tessellated surface<sup>45</sup>. Average cortical thickness  
158 estimates were obtained for 358 brain regions based on the Glasser anatomical atlas<sup>56</sup>.

159 **Task-based functional MRI.** Preprocessing was performed using SPM8  
160 ([www.fil.ion.ucl.ac.uk/spm](http://www.fil.ion.ucl.ac.uk/spm)) and included slice time correction, realignment to the first  
161 image, coregistration with the structural image, spatial normalization into MNI space and

162 smoothing (8 mm Gaussian kernel). Details of the paradigm and the statistical first-level  
163 model have been provided elsewhere<sup>30</sup>. In brief, the paradigm involved a probabilistic value-  
164 based decision making task<sup>57</sup> in which participants could receive a fixed monetary reward in  
165 each of 201 trials. While the goal of the paradigm is to distinguish model-based from model-  
166 free learning signals by means of a computational model, following our rationale of  
167 constructing a neuroimaging battery with only the most fundamental modalities, we here used  
168 the basic contrast of reward (+20 Euro cents) versus no reward.

169 Within the statistical first-level model<sup>30</sup>, we used the two onset regressors for *reward* and *no*  
170 *reward* conditions during the outcome phase and computed directional contrasts (*reward* > *no*  
171 *reward*). To reduce the extremely high dimensionality of individual fMRI whole-brain  
172 contrast maps (>200.000 valid voxels), we downsampled the data to 10 mm isotropic voxels  
173 resulting in 1461 valid voxels (features) for each participant. Exploratory classification  
174 analyses with smaller voxels sizes (2 - 8 mm in steps of 2 mm) showed that no relevant  
175 information was lost at 10 mm.

176 **Resting-state functional MRI.** Preprocessing of resting-state data was performed using FSL  
177 (<https://www.fmrib.ox.ac.uk/fsl>) and included motion correction, slice timing correction, non-  
178 brain removal, 6 mm FWHM spatial smoothing. Data were denoised using independent  
179 component analysis (ICA-AROMA<sup>58,59</sup>), high-pass-filtered at 0.008 Hz and normalized to  
180 MNI space. To estimate whole-brain accumbens connectivity maps, binary seed masks of the  
181 left and right nucleus accumbens were defined using the Harvard-Oxford Subcortical  
182 Probability Atlas (50% probability threshold). The first Eigen time series was extracted from  
183 the preprocessed resting-state data, separately for the left and right nucleus accumbens.  
184 Finally, these time series were regressed against every other voxel's time series using deep  
185 white matter and CSF as nuisance variables. To reduce dimensionality, we computed regional  
186 averages for the 110 anatomical regions of the JHU brain atlas<sup>43</sup> and averaged across the left  
187 and right nucleus accumbens. Note that while participants were excluded when falling asleep  
188 during scanning, part of the interindividual variance in resting state connectivity may be  
189 explained by unknown interindividual differences in drowsiness/vigilance, which were not  
190 assessed in an objective manner in the current study. Nevertheless, only a small number of  
191 participants (7 AUD and 6 HC) reported brief moments of drowsiness or loss of vigilance  
192 during the exit interview.

193 **Unimodal diagnostic classification**

194 In a first step, the diagnostic accuracy of each of the five modalities was assessed. Before  
195 submitting the modalities to classification, each modality was soft-normalized by removing  
196 the median and scaling the data according to a quantile range (1% to 99%)<sup>60</sup>. We considered  
197 two different classifier types: support vector machine (SVM<sup>61</sup>) and weighted robust distance  
198 (WeiRD<sup>16,62,63</sup>). SVM is arguably the most popular classifier in the context of  
199 neuroimaging<sup>64</sup>, as it robustly handles data with high dimensionality but few samples per  
200 class. We used the implementation provided by libsvm<sup>65</sup> and the default kernel (radial basis  
201 function). WeiRD is a distance-to-centroid classifier (code available at <https://github.com/m-guggenmos/weird>), which assigns class labels based on the Manhattan distance to class  
202 prototypes in a statistically weighted feature space. A key advantage of WeiRD is the fact  
203 that it is parameter-free and hence does not require expensive optimization schemes. In  
204 previous work we have shown that WeiRD, despite its simplicity, performs surprisingly well  
205 across a range of neuroimaging classification problems<sup>16,62,63</sup>.

207 Both classifiers operated in a leave-one-out (LOO) cross-validation procedure, such that in  
208 each cross-validation fold, the classifiers were trained on the data of all but one participant  
209 and tested on the left-out participant. Each participant was left out and predicted exactly once.  
210 LOO was chosen over other cross-validation schemes, because it is approximately  
211 unbiased<sup>66</sup>, deterministic and computationally inexpensive. For SVM specifically, we  
212 additionally performed nested cross-validation and grid search to optimize the cost parameter  
213  $C$  (range of  $C: 2^x, x = -5:1:10$ ). We used the balanced accuracy as a scoring metric and  
214 computed p-values of the null hypothesis that the balanced accuracy is not above chance<sup>67</sup>.

215 Note that the number of features outnumbered the number of samples. This is the standard  
216 scenario in multivariate neuroimaging analyses, which have been widely used with success to  
217 fit machine learning models within and across individuals<sup>68,69</sup>, most likely due to  
218 intercorrelations between features<sup>70</sup>.

219 In addition, we assessed the importance of each region for classification. For clarity and  
220 brevity, we show importance scores only for the superior classifier (SVM or WeiRD) of a  
221 modality. Where SVM was superior, we derived importance scores from the hyperplane-  
222 defining weight vector in feature space. Although the feature space is not the input space in  
223 the case of the radial-basis function kernel used here, the weights in feature space  
224 nevertheless provided a useful estimate of feature importance for descriptive purposes. Where

225 WeiRD was superior, we used WeiRD votes as a measure of feature importance, as described  
226 previously<sup>16</sup>.

227 **Multimodal diagnostic classification**

228 The aim of multimodal neuroimaging classification is to improve overall performance by  
229 leveraging on the combined information of more than one neuroimaging modality. Here we  
230 combined all five modalities introduced above (GMD, CSF, CTH, RWR, NAC) in an  
231 integrated classification scheme. The key design principle for the multimodal classification  
232 scheme was that a separate classifier was trained on each modality, which enabled us to both  
233 select the best classifier for each modality and to fit it to the specific statistical properties of a  
234 modality.

235 As for unimodal classification, multimodal classification was based on soft-normalised data  
236 and proceeded in a LOO cross-validation procedure. In the multimodal scenario this meant  
237 that in each cross-validation fold, all data (i.e. all modalities) of one participant were left out  
238 for independent testing. Henceforth, we describe the analytic work-flow of training and  
239 prediction for one such fold (see also Figure 1).

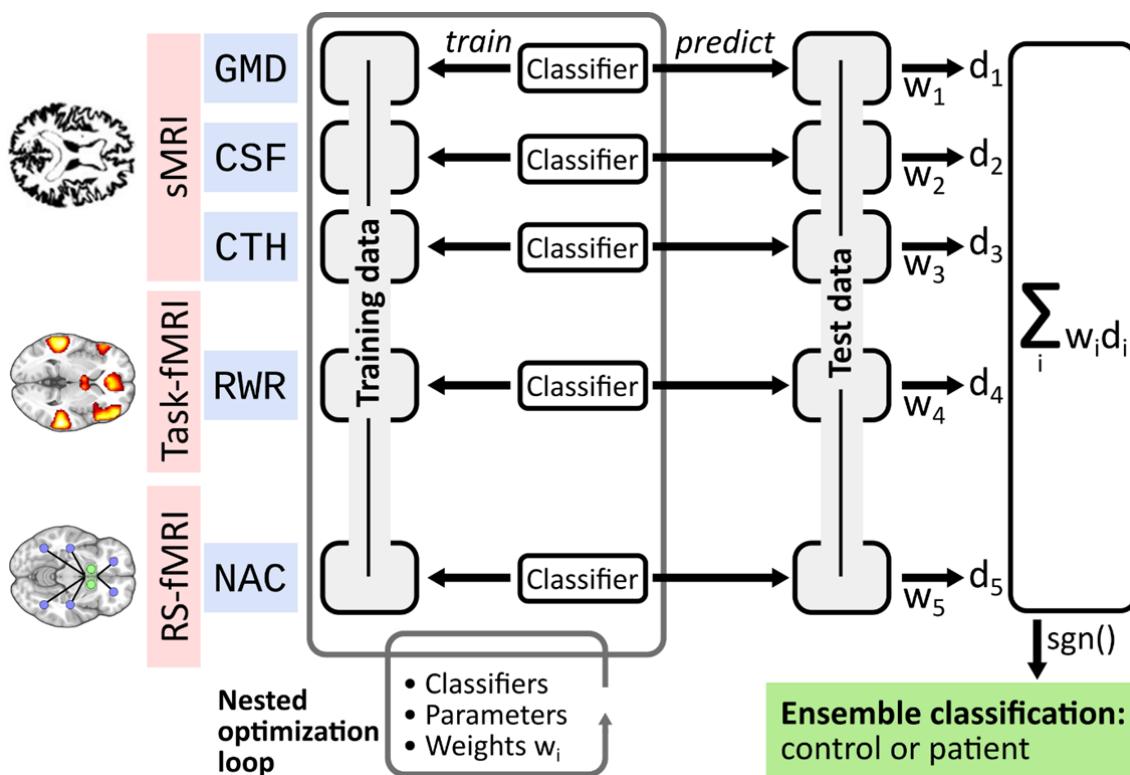
240 **Training.** During training, a separate classifier was trained on each modality and three  
241 factors were optimized in a nested cross-validation procedure: the classifier type (SVM or  
242 WeiRD); the parameters of a classifier if applicable (cost parameter C for SVM); and  
243 modality-specific weights. In detail, for each modality separately it was first assessed whether  
244 SVM (with second-level-nested cross-validation for the optimization of C) or WeiRD were  
245 more accurate on the training data set. If SVM was superior, the optimal cost parameter was  
246 then estimated anew on the entire training data. After determining the optimal classifier for  
247 each modality, a new nested cross-validation procedure was performed for all modalities  
248 combined to estimate weighting factors for each modality. To limit computational complexity  
249 and increase robustness, each of the five modalities could be weighted only with a factor of  
250 either 1 or 2.

251 **Prediction.** After training, the classification scheme was applied to the yet unseen test  
252 subject. Each modality-specific classifier computed a signed continuous decision value for  
253 the respective modality, where negative and positive values represented control and patient  
254 predictions and the absolute value the certainty of a classifier. The overall ensemble

255 prediction was based on the sign of a weighted sum of modality-specific decision values with  
256 weighting factors determined during training.

257 To assess the effect of optimizing the classifier and the weighing scheme, the entire  
258 procedure was performed with either fixing the classifier to either WeiRD or SVM (but still  
259 optimizing C), and/or omitting modality weighting, i.e. fixing all weights to 1.

260 Finally, note that a key advantage of our classification scheme was that missing data were of  
261 no concern: if data were not available for a single modality (see Table 2 for the number of  
262 valid samples in each modality), it was omitted from the weighted sum of modality-specific  
263 decision values.



265 **Figure 1. Multimodal classification scheme.** Depicted is one exemplary split into training data and test data. Using a nested  
266 optimization loop, three modality-specific factors are optimized on the training data: classifier types (SVM, WeiRD),  
267 parameters (cost parameter  $C$  for SVM) and weights  $w_i$ . The trained and optimized model is then applied to the test data and  
268 continuous decision values  $d_i$  are computed for each modality-specific classifier. The final diagnostic classification is based  
269 on a weighted sum of decision values, where weights correspond to those estimated during training.

## 270 Data availability

271 Code and data used in the current study are available from the corresponding author on  
272 reasonable request.

273 **RESULTS**

274 **1. Unimodal diagnostic classification**

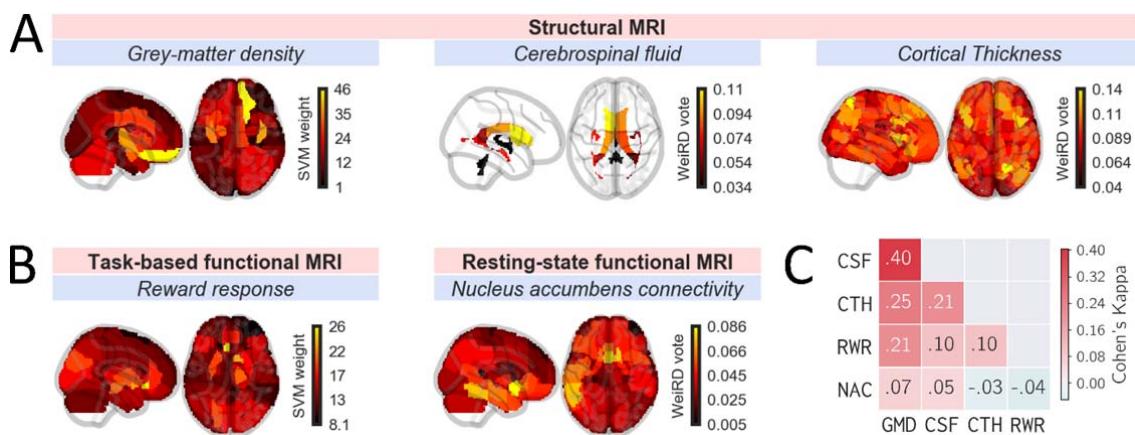
275 Three reasons motivated an initial assessment of unimodal classification, i.e. classification  
276 based on each modality individually. First, to provide a reference against which multimodal  
277 classification could be benchmarked. Second, to elucidate for each modality which brain  
278 regions contained information predictive of diagnosis. And third, to assess the agreement  
279 between different modalities with respect to patient/control predictions (inter-modality  
280 reliability).

281 We found that each modality discriminated significantly between patients and controls  
282 (Table 3), although the accuracy varied strongly across modalities. For both SVM and  
283 WeiRD, grey-matter density was the best-performing modality with balanced accuracies of  
284 76.6% and 71.3%, respectively. All other modalities ranged between 55 and 66% accuracy.  
285 The balance between specificity and sensitivity was heterogenous across modalities. For  
286 instance, while cortical thickness was more sensitive than specific, cerebrospinal fluid and  
287 reward responses showed an inverse pattern.

288 To investigate which brain regions contributed to classification, we inspected feature  
289 importances for each modality (Figure 2A/B and Supplementary Figure S1). Across  
290 modalities, we found that anterior cingulate and inferior frontal brain regions discriminated  
291 best between patients and controls. This included the cerebrospinal fluid, for which the (left)  
292 frontal ventricle was most discriminative. These results are broadly in line with the  
293 hypothesis of prefrontal cortex dysfunction in addiction<sup>71-73</sup>. In addition, for functional  
294 reward responses the nucleus accumbens was the second most discriminative brain region  
295 (after subcallosal anterior cingulate cortex), a key region of generic reward pathways<sup>74</sup> and of  
296 bottom-up theories of addiction such as the incentive-sensitization theory<sup>22</sup>. Although not the  
297 focus of this study, the brain regions that discriminate best between patients and controls map  
298 well on established neural correlates of alcohol dependence<sup>75,76</sup> and provide validation to the  
299 approach taken in this study.

300 Our initial unimodal analysis of feature importances showed that informative features were  
301 primarily located in (orbito-)prefrontal and cingulate brain regions as well as in the nucleus  
302 accumbens (for basic reward responses).

303 How well do different modalities agree with respect to their diagnostic predictions? To find  
 304 out, we computed inter-modality reliability scores based on Cohen's Kappa<sup>77</sup>, which  
 305 measures the agreement of two "raters" over and above the agreement expected by chance.  
 306 For each modality we used the classifier (SVM or WeiRD) that performed better for a given  
 307 modality. As shown in Figure 2C, the agreement was generally highest between modalities  
 308 based on structural MRI. Within those, predictions based on grey-matter density and  
 309 cerebrospinal fluid showed the highest agreement with a Kappa value of 0.4 (considered a  
 310 'moderate' agreement<sup>78</sup>). All other modality comparisons exhibited relatively little agreement  
 311 with Kappa values <=0.25. Across all modalities, predictions based on nucleus accumbens  
 312 connectivity showed the least agreement with any other modality.



314 **Figure 2. Unimodal classification.** Feature importances of (A) structural and (B) functional neuroimaging modalities.  
 315 Depicted are 2-d projections ('glass brains') of feature importances along the x- and z-axis. Feature importances represent  
 316 SVM weights (grey-matter density, reward response) or WeiRD votes (cerebrospinal fluid, cortical thickness, resting state)  
 317 depending on which classifiers was superior for a given modality. (C) Inter-modality reliability matrix based on Cohen's  
 318 Kappa describing the diagnostic agreement between modalities.

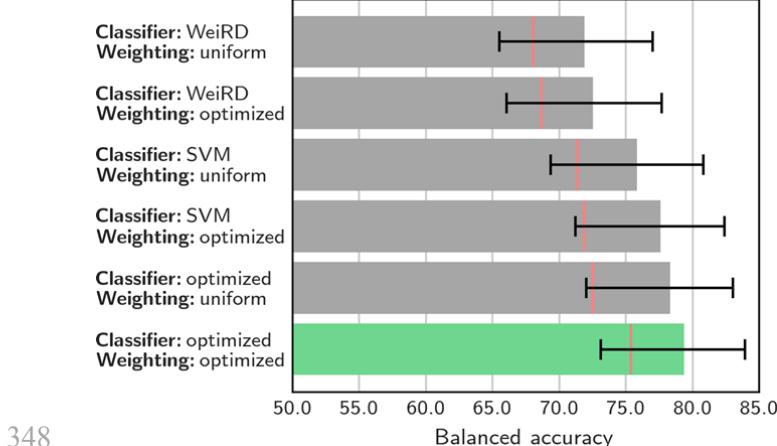
## 319 **2. Multimodal diagnostic classification performs better than each individual 320 modality**

321 To combine modalities for multimodal classification, we trained individual classifiers on each  
 322 modality separately, computed predictions based on each modality, and then combined these  
 323 predictions in the form of a weighted ensemble vote (Figure 1). A number of details of this  
 324 analytic framework deserve highlighting. First, classifiers were optimized for each modality  
 325 individually, both with respect to the classifier type (SVM or WeiRD) and in terms of a  
 326 modality-specific regularization parameter (C in the case of SVM). Second, instead of relying  
 327 on discrete predictions, continuous decision values were computed for each modality, which  
 328 enabled a more fine-grained ensemble decision. Third, on top of these intrinsic classifier-

329 based decision values, an external weighting matrix across modalities was learned from the  
330 training data, further refining the ensemble vote.

331 Employing this fully-featured multimodal classification scheme yielded a balanced accuracy  
332 of 79.3% ( $p < 10^{-18}$ ). Thus, combining the information from multiple modalities improved  
333 the accuracy compared to the best individual modality (grey-matter density: 76.6%). An  
334 analysis of sensitivity and specificity showed that this improvement was due to an increase of  
335 specificity (81.3%; sensitivity: 77.3%), which reached a level unmatched by any individual  
336 modality.

337 To investigate which analytic component of the multimodal classification scheme was  
338 essential for this improvement – or which aspect may have been redundant – we computed  
339 accuracies while muting one or more of the features. As shown in Figure 3, we found that  
340 indeed each analytic component discussed above (optimization of classifiers, decision values  
341 instead of discrete predictions, cross-modal weighting) was responsible for an incremental  
342 improvement of performance. Of these, the largest improvement was due to considering  
343 continuous decision values instead of binary “control” and “patient” predictions (represented  
344 by red lines in Figure 3): without this feature, the multimodal accuracy would have dropped  
345 below the best individual modality, even if classifier and weights were optimized (75.4%).  
346 Overall, these results show that multimodal classification requires a nuanced integration of  
347 modalities in order to achieve a meaningful benefit.



348

349 **Figure 3. Multimodal classification.** Balanced accuracy for classification schemes based on different classifier  
 350 configurations (SVM, WeiRD or optimized between SVM and WeiRD) and with uniform (i.e., all weights set to 1) or  
 351 optimized weighting of modalities. Optimizing both classifiers and weighting yielded the best performance (highlighted in  
 352 green). Red lines indicate the balanced accuracy when ensemble prediction was based on discrete “control” and “patient”  
 353 judgements instead of continuous decision values. Error bars represent the 95% posterior probability interval<sup>67</sup>.

354 To assess the significance of each individual modality for multimodal classification, we used  
 355 the fully-featured classification scheme and computed classification accuracies while  
 356 tentatively excluding each modality once. The results (Table 4) showed that the accuracy  
 357 dropped in each case, indicating that each modality was important for overall multimodal  
 358 performance. Not surprisingly, the sharpest drop in performance was observed when  
 359 excluding the best individual modality, grey-matter density (−9.6%). In particular, grey-  
 360 matter density was largely responsible for the high sensitivity of multimodal classification,  
 361 evidenced by a substantial drop of sensitivity when excluded (−18.1%). By contrast, the  
 362 sharpest drop in specificity was found when excluding cerebrospinal fluid (−7.3%). Taken  
 363 together, these results show that the success of multimodal classification was based on the  
 364 combined information of all modalities such that no single modality was effectively  
 365 redundant.

## 366 DISCUSSION

367 In the present investigation we used structural, functional task-based and functional  
 368 resting-state MRI to construct a diagnostic machine-learning classifier for alcohol  
 369 dependence. A novel multimodal classification scheme, in which modality-specific  
 370 weightings and optimal classifiers were estimated from training data, slightly outperformed  
 371 the strongest individual modality and achieved a balanced accuracy of 79.3%.

372 Our initial unimodal analysis of feature importances showed that informative features were  
 373 primarily located in (orbito-)prefrontal and cingulate brain regions as well as in the nucleus

374 accumbens (for basic reward responses). These foci are broadly in line with previous  
375 investigations into brain structure and function in alcohol dependence, which identified  
376 executive and reward networks as major neural circuits that are affected in patients<sup>75</sup>. Thus,  
377 our unimodal structural and functional data show diagnostic characteristics well in line with  
378 previous results and should replicate reliably in future studies. Despite the overlap of affected  
379 brain structures, none of these considered modalities was redundant: multimodal  
380 classification accuracy was reduced for each modality that was tentatively excluded.  
381 Together, these exploratory analyses into individual modalities thus attest to 1) plausible  
382 unimodal between-group effects and 2) a sensible selection of modalities that leveraged on  
383 non-redundant sources of information.

384 In comparison to the best individual modality (grey-matter density with SVM: 76.6%),  
385 multimodal classification yielded an improvement of 2.7% accuracy, corresponding to 6  
386 additional subjects that were correctly classified. The gain in accuracy was specifically due to  
387 an increase in specificity (81.3%), which was 7% higher than for grey-matter density  
388 (74.2%). The specific increase of specificity is noteworthy, as we recently found a particular  
389 advantage of unimodal computer-based classification over human judgements with respect to  
390 *sensitivity*, while *specificity* was higher for the judgements of an experienced radiologist  
391 (81.4%<sup>16</sup>). Thus, when combining multiple modalities, computer-based classification  
392 matched human performance in terms of identifying true negatives, thereby eliminating a  
393 weakness of unimodal classification.

394 Although the improvement in overall accuracy is modest at 2.7%, one important qualification  
395 is that the comparison to the best of several modalities is necessarily unfair due to selection  
396 bias; that is, it is likely that the accuracy of the best modality is inflated by noise. Besides this  
397 ‘double dipping’<sup>79</sup> bias, even at 2.7% our observed benefit of multimodal classification is  
398 noteworthy, as many previous studies have failed to find any advantage of combining  
399 multiple modalities<sup>9,10,80,81</sup>. There is currently no consensus on *why* an effective combination  
400 of multiple MRI modalities appears to be a surprisingly intricate task. But clearly, success of  
401 multimodal classification depends on how individual modalities complement each other. If  
402 individual modalities perfectly agree with each other, they are effectively redundant and thus  
403 no multimodal benefit is expected. Vice versa, if predictions of individual modalities are  
404 completely uncorrelated, there is no positive cumulative effect through the combination of  
405 modalities. In the present case, the agreement of predictions between different modalities was

406 quite variable overall but tended towards low inter-modality reliabilities (Cohen's Kappa  
407 scores up to 0.4), which may explain the moderate benefit of multimodal classification.

408 At a methodological level, we found a number of analytic choices to be critical for the  
409 integration of multiple modalities. First, as data from different modalities will naturally have  
410 different structural properties, it is likely that there is no single classifier type that fits all  
411 modalities. Yet, to our knowledge, while multimodal investigations often *compare* classifier  
412 types<sup>9,11,82</sup>, they do so by applying the same classifier to all modalities, i.e. they do not  
413 optimize the classifier type in a modality-specific manner. Here we found that optimizing the  
414 classifier type for each modality was superior to using either SVM or WeiRD uniformly  
415 across modalities. Second, applying a weight matrix learned from training data to the  
416 ensemble prediction likewise increased the performance compared to unweighted integration.  
417 Although this particular aspect has been considered previously (e.g. by means of logistic  
418 regression on predictions of individual modalities<sup>11</sup>), our results corroborate the importance  
419 of this analytic step.

420 Third, the biggest gain in accuracy was owed to using continuous classifier decision values  
421 instead of discrete (binary) predictions. The benefit of decision values can be explained by  
422 the fact that for both SVM and WeiRD, decision values reflected the certainty of the  
423 classifier. Thus, more certain predictions factor in more strongly into the ensemble prediction  
424 and thereby improve the overall accuracy. Indeed, we have previously shown that valuable  
425 information is contained in unthresholded decision values of classifiers applied to  
426 neuroimaging data<sup>63</sup>. On the basis of our results we thus recommend to avoid discretizing  
427 modality-specific predictions and instead to utilize fine-grained information contained in  
428 classifier decision values.

429 Despite these methodological insights about combining multiple neuroimaging modalities  
430 from both structural and functional MRI, considering the fact that fMRI scanning was more  
431 time-consuming and involved more elaborate data analysis, the achieved improvement  
432 through fMRI must be debated in view of clinical practicability and a cost-benefit analysis. In  
433 our view, several factors are important for fMRI to become part of a clinically realistic  
434 diagnostic imaging battery. First, an accuracy net gain in the order of a few percent appears  
435 relatively small, given the additional effort of instructing and conducting a functional scan in  
436 addition to a structural scan. Thus, the task-based and resting-state fMRI signals employed by  
437 a multimodal approach should preferably either feature more sensitive measures – or

measures that are even more orthogonal to the information from structural MRI. Second, employed fMRI tasks have to be sufficiently brief. This would be a relatively simple optimization of the present imaging battery, as the basic reward signal used for classification was extracted from a relatively complex decision-making task which could be condensed substantially. Third, at present the analysis of functional MRI data is more laborious compared to the analysis of structural MRI data. However, with the development of standardized, efficient, and robust analysis protocols (e.g. fMRIprep<sup>83</sup>), fMRI could become a realistic option for day-to-day clinical diagnosis. In sum, in terms of direct clinical applicability, currently the most realistic neuroimaging-based classifier for AD may be unimodal based on structural MRI and grey-matter density specifically.

A number of limitations should be noted. First, our sample is predominantly male, which limits the generalization of our results to female patients. This may be especially relevant for the most predictive modality in our approach, grey-matter density, as numerous studies have shown greater sensitivity to the neurotoxic effects of alcohol on grey matter in women<sup>84-86</sup>. Second, we were not able to validate our results against an independent sample, as the multimodal imaging battery employed by the LeAD study has no precedence in the AD literature. However, in previous work on unimodal imaging<sup>16</sup> we found almost perfect generalization to an independent sample (original: 74%; generalization: 73%). Although these previous results cannot be directly extrapolated to our current work, we note that both investigations largely share the robust methodology, including the use of the weighted robust distance classifier<sup>62</sup> and rigorous cross-validation with nested cross-validation for tuning parameters.

Third, the selection of neuroimaging modalities is not exhaustive. While in the present work we focused on a parsimonious set of established but basic modalities, especially in the functional domain there is abundant literature on various functional correlates of causes, state markers (e.g. craving) and consequences in alcohol dependence<sup>75,87</sup>. For instance, the influential iRISA (impaired response inhibition and salience attribution) model proposes that disrupted function of the prefrontal cortex leads not only to attributing excessive salience to drugs and associated cues, but also impairs the ability to inhibit drug-related behaviours<sup>71</sup>. More complete neuroimaging models of alcohol dependence could thus additionally consider a prefrontal functional correlate of cognitive control. Regarding functional connectivity, there is evidence that large-scale functional networks (e.g. default mode network or cognitive control network) explain a substantial amount of variance with respect to alcohol use severity

471 and may as well be informative for diagnostic classification<sup>88</sup>. These large-scale networks  
472 were not considered in the present study and thus deserve further research in the context of  
473 diagnostic neuroimaging-based classification. Fourth, based on the cross-sectional design in  
474 this study one cannot infer whether the neurobiological differences utilized by our classifier  
475 are causes, state markers or consequences of alcohol dependence (although we provide  
476 preliminary evidence for the latter possibility in supplementary Fig. 2). Thus, our classifier  
477 provides no mechanistic insight into the pathogenesis of alcohol dependence. For instance,  
478 based on previous findings it is possible to that group differences in brain structure may be  
479 predominantly consequences of severe alcohol abuse, as these changes partially reverse  
480 during abstinence<sup>89–91</sup>; differences in brain function may, in turn, predispose for addiction or  
481 relapse from addiction<sup>30,92</sup>. From a strict machine-learning point of view, one may be tempted  
482 to treat the issue of mechanistic insight as secondary. However, although neural features that  
483 characterize a predisposition are particularly valuable due to their prognostic potential, for  
484 diagnosis they bear the risk of misclassifying healthy individuals that did not develop AD  
485 despite having a disadvantageous predisposition. Patients which developed AD without such  
486 a predisposition could likewise be misclassified due to these features. For this reason, the  
487 ‘chicken or egg’ causality dilemma is indeed relevant for machine-learning-based diagnosis  
488 and should be investigated in future longitudinal studies.

489 Finally, in view of modern frameworks of psychiatric disease such as the Research Domain  
490 Criteria (RDoC) project, in which psychiatric phenotypes are defined as “spanning the range  
491 from normal to abnormal”<sup>93,94</sup>, the present approach could be readily adapted to predict  
492 dimensional markers of disease. Analogous to the idea of a weighted voting scheme across  
493 classifiers applied here for the case of binary prediction, multiple modality-specific  
494 regression models would be trained and their outputs combined to form a continuous  
495 ensemble prediction. The ensemble prediction would likewise be based on a weighted (and  
496 normalized) sum of individual predictions. Conceivable dimensional markers for the case of  
497 alcohol dependence are the magnitude of craving in acute addiction<sup>95</sup>, biological markers  
498 such as serum levels of carbohydrate deficient transferrin (CDT) and gamma  
499 glutamyltransferase (GGT)<sup>96</sup>, or scores of clinical questionnaires such as the Alcohol Use  
500 Disorders Identification Test (AUDIT)<sup>97</sup>. Overall, we conclude that the combination of  
501 multiple neuroimaging modalities is able to moderately improve the accuracy of machine-  
502 learning-based diagnostic classification in alcohol dependence. Our results allow us to make  
503 several methodological recommendations for the exploitation and integration of different

504 modalities with the goal to compute optimal ensemble predictions, thereby paving the way  
505 towards more effective multimodal neuroimaging classifiers. Yet, at present, given the strong  
506 predictive performance of grey-matter density alone and taking a cost-benefit analyses into  
507 account, we currently recommend to focus on structural MRI for the diagnostic classification  
508 of alcohol dependence.

509

510 **References**

- 511 1. Liu, S. *et al.* Multimodal neuroimaging computing: a review of the applications in neuropsychiatric  
512 disorders. *Brain Informatics* **2**, 167–180 (2015).
- 513 2. Perrin, R. J., Fagan, A. M. & Holtzman, D. M. Multimodal techniques for diagnosis and prognosis of  
514 Alzheimer's disease. *Nature* **461**, 916–922 (2009).
- 515 3. Calhoun, V. D. & Sui, J. Multimodal Fusion of Brain Imaging Data: A Key to Finding the Missing  
516 Link(s) in Complex Mental Illness. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **1**, 230–244 (2016).
- 517 4. Santelmann, H., Franklin, J., Bußhoff, J. & Baethge, C. Test-retest reliability of schizoaffective disorder  
518 compared with schizophrenia, bipolar disorder, and unipolar depression-a systematic review and meta-  
519 analysis. *Bipolar Disord.* **17**, 753–768 (2015).
- 520 5. Freedman, R. *et al.* The initial field trials of DSM-5: New blooms and old thorns. *Am. J. Psychiatry* **170**,  
521 1–5 (2013).
- 522 6. Wardenaar, K. J. & de Jonge, P. Diagnostic heterogeneity in psychiatry: Towards an empirical solution.  
523 *BMC Med.* **11**, 2–4 (2013).
- 524 7. Jablensky, A. Psychiatric classifications: Validity and utility. *World Psychiatry* **15**, 26–31 (2016).
- 525 8. Callard, F. Psychiatric diagnosis: The indispensability of ambivalence. *J. Med. Ethics* **40**, 526–530  
526 (2014).
- 527 9. Ritter, K. *et al.* Multimodal prediction of conversion to Alzheimer based on incomplete  
528 biomarkers\*This work was supported by the Bernstein Computational Program of the German Federal  
529 Ministry of Education and Research (01GQ1001C, 01GQ0851, GRK 1589/1), the European Regiona.  
530 *Alzheimer's Dement. Diagnosis, Assess. Dis. Monit.* **1**, 206–215 (2015).
- 531 10. Shaffer, J. L. *et al.* Predicting Cognitive Decline in Subjects at Risk for Alzheimer Disease by Using  
532 Combined. *Radiology* **266**, 583–591 (2013).
- 533 11. Young, J. *et al.* Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in  
534 patients with mild cognitive impairment. *NeuroImage Clin.* **2**, 735–745 (2013).
- 535 12. Klöppel, S. *et al.* Automatic classification of MR scans in Alzheimer's disease. *Brain* **131**, 681–689  
536 (2008).
- 537 13. Hackmack, K. *et al.* Can we overcome the 'clinico-radiological paradox' in multiple sclerosis? *J.*  
538 *Neurol.* **259**, 2151–2160 (2012).
- 539 14. Weygandt, M. *et al.* MRI pattern recognition in multiple sclerosis normal-appearing brain areas. *PLoS*  
540 *One* **6**, (2011).
- 541 15. Harper, C. & Matsumoto, I. Ethanol and brain damage. *Curr. Opin. Pharmacol.* **5**, 73–78 (2005).
- 542 16. Guggenmos, M. *et al.* Decoding diagnosis and lifetime consumption in alcohol dependence from grey-  
543 matter pattern information. *Acta Psychiatr. Scand.* **137**, 252–262 (2018).
- 544 17. Jernigan, T. L. *et al.* Reduced cerebral grey matter observed in alcoholics using magnetic resonance  
545 imaging. *Alcohol. Clin. Exp. Res.* **15**, 418–427 (1991).
- 546 18. Pfefferbaum, A. *et al.* Brain gray and white matter volume loss accelerates with aging in chronic  
547 alcoholics: a quantitative MRI study. *Alcohol. Clin. Experimental Res.* **16**, 1078–89 (1992).
- 548 19. Fox, J. H., Ramsey, R. G., Huckman, M. S. & Proske, A. E. Cerebral Ventricular Enlargement Chronic.

- 549                   *JAMA* **4**, 365–368 (1976).
- 550       20. Ron, M. A. *The alcoholic brain: CT scan and psychological findings. Psychological Medicine. Monograph Supplement* **3**, (1983).
- 551       21. Kraemer, G. W., Raymond Lake, C., Ebert, M. H. & McKinney, W. T. Effects of alcohol on cerebrospinal fluid norepinephrine in rhesus monkeys. *Psychopharmacology (Berl)*. **85**, 444–448 (1985).
- 552       22. Robinson, T. E. & Berridge, K. C. The neural basis of drug craving: An incentive-sensitization theory of addiction. *Brain Res. Rev.* **18**, 247–291 (1993).
- 553       23. Wrage, J. *et al.* Dysfunction of reward processing correlates with alcohol craving in detoxified alcoholics. *Neuroimage* **35**, 787–794 (2007).
- 554       24. Alba-Ferrara, L., Müller-Oehring, E. M., Sullivan, E. V., Pfefferbaum, A. & Schulte, T. Brain responses to emotional salience and reward in alcohol use disorder. *Brain Imaging Behav.* **10**, 136–146 (2016).
- 555       25. Tapert, S. F. *et al.* Neural response to alcohol stimuli in adolescents with alcohol use disorder. **60**, 727–735 (2003).
- 556       26. Huang, M. C. *et al.* Prevalence and identification of alcohol use disorders among severe mental illness inpatients in Taiwan. *Psychiatry Clin. Neurosci.* **63**, 94–100 (2009).
- 557       27. Ruskin, P. *et al.* Reliability and acceptability of psychiatric diagnosis via telecommunication and audiovisual technology. *Psychiatr. Serv.* **49**, 1086–1088 (1998).
- 558       28. Martin, C. S., Pollock, N. K., Bukstein, O. G. & Lynch, K. G. Inter-rater reliability of the SCID alcohol and substance use disorders section among adolescents. *Drug Alcohol Depend.* **59**, 173–176 (2000).
- 559       29. Guggenmos, M. *et al.* Quantitative neurobiological evidence for accelerated brain aging in alcohol dependence. *Transl. Psychiatry* **7**, 1279 (2017).
- 560       30. Sebold, M. *et al.* When habits are dangerous - Alcohol expectancies and habitual decision-making predict relapse in alcohol dependence. *Biol. Psychiatry* **82**, 847–856 (2017).
- 561       31. Garbusow, M. *et al.* Pavlovian-to-instrumental transfer effects in the nucleus accumbens relate to relapse in alcohol dependence. *Addict. Biol.* **21**, 719–731 (2016).
- 562       32. Sebold, M. *et al.* Don't Think, Just Feel the Music: Individuals with Strong Pavlovian-to-Instrumental Transfer Effects Rely Less on Model-based Reinforcement Learning. *J. Cogn. Neurosci.* **28**, 985–995 (2016).
- 563       33. Veer, I. *et al.* Nucleus Accumbens Functional Connectivity at Rest is Related to Alcohol Consumption in Young Adults. in *Society of Biological Psychiatry 72nd Annual Scientific Convention and Meeting* **81**, S61 (Elsevier, 2017).
- 564       34. Garbusow, M. *et al.* Pavlovian-to-Instrumental Transfer in Alcohol Dependence: A Pilot Study. *Neuropsychobiology* **70**, 111–121 (2014).
- 565       35. Garbusow, M., Sebold, M., Beck, A. & Heinz, A. Too Difficult to Stop: Mechanisms Facilitating Relapse in Alcohol Dependence. *Neuropsychobiology* **70**, 103–110 (2014).
- 566       36. Wittchen, H.-U. & Pfister, H. *DIA-X-Interviews: Manual für Screening-Verfahren und Interview; Interviewheft.* (Swets & Zeitlinger, 1997).
- 567       37. Jacobi, F. *et al.* The design and methods of the mental health module in the German Health Interview and Examination Survey for Adults (DEGS1-MH). *Int. J. Methods Psychiatr. Res.* **22**, 83–99 (2013).

- 589 38. Oldfield, R. C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*  
590 **9**, 97–113 (1971).
- 591 39. Kotz, D., Böckmann, M. & Kastaun, S. Nutzung von Tabak und E-Zigaretten sowie Methoden zur  
592 Tabakentwöhnung in Deutschland. *Dtsch. Arztbl. Int.* **115**, 235–242 (2018).
- 593 40. Fortier, C. B. *et al.* Reduced Cortical Thickness in Abstinent Alcoholics and Association with Alcoholic  
594 Behavior. *Alcohol. Clin. Exp. Res.* **35**, 2193–2201 (2011).
- 595 41. Mashhoon, Y. *et al.* Binge alcohol consumption in emerging adults: Anterior cingulate cortical  
596 ‘thinness’ is associated with alcohol use patterns. *Alcohol. Clin. Exp. Res.* **38**, 1955–1964 (2014).
- 597 42. Momenan, R. *et al.* Effects of alcohol dependence on cortical thickness as determined by magnetic  
598 resonance imaging. *Psychiatry Res. - Neuroimaging* **204**, 101–111 (2012).
- 599 43. Faria, A. V *et al.* Atlas-based analysis of resting-state functional connectivity: Evaluation for  
600 reproducibility and multi-modal anatomy–function correlation studies. *Neuroimage* **61**, 613–621 (2012).
- 601 44. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical Surface-Based Analysis. I. Segmentation and Surface  
602 Reconstruction Anders. *Neuroimage* **9**, 179–194 (1999).
- 603 45. Fischl, B. & Dale, A. M. Measuring the thickness of the human cerebral cortex from magnetic  
604 resonance images. *Proc. Natl. Acad. Sci.* **97**, 11050–11055 (2000).
- 605 46. Fischl, B., Sereno, M. I. & Dale, A. M. Cortical Surface-Based Analysis. II: Inflation, Flattening, and a  
606 Surface-Based Coordinate System. *Neuroimage* **9**, 195–207 (1999).
- 607 47. Fischl, B., Sereno, M. I., Tootell, R. B. H. & Dale, A. M. High-resolution inter-subject averaging and a  
608 surface-based coordinate system. *Hum. Brain Mapp.* **8**, 272–284 (1999).
- 609 48. Fischl, B., Liu, A. & Dale, A. M. Automated manifold surgery: Constructing geometrically accurate and  
610 topologically correct models of the human cerebral cortex. *IEEE Trans. Med. Imaging* **20**, 70–80  
611 (2001).
- 612 49. Fischl, B. *et al.* Whole brain segmentation: Automated labeling of neuroanatomical structures in the  
613 human brain. *Neuron* **33**, 341–355 (2002).
- 614 50. Fischl, B. *et al.* Automatically Parcellating the Human Cerebral Cortex. *Cereb. Cortex* **14**, 11–22  
615 (2004).
- 616 51. Fischl, B. *et al.* Sequence-independent segmentation of magnetic resonance images. *Neuroimage* **23**,  
617 69–84 (2004).
- 618 52. Han, X. *et al.* Reliability of MRI-derived measurements of human cerebral cortical thickness: The  
619 effects of field strength, scanner upgrade and manufacturer. *Neuroimage* **32**, 180–194 (2006).
- 620 53. Jovicich, J. *et al.* Reliability in multi-site structural MRI studies: Effects of gradient non-linearity  
621 correction on phantom and human data. *Neuroimage* **30**, 436–443 (2006).
- 622 54. Ségonne, F. *et al.* A hybrid approach to the skull stripping problem in MRI. *Neuroimage* **22**, 1060–1075  
623 (2004).
- 624 55. Reuter, M., Schmansky, N. J., Rosas, H. D. & Fischl, B. Within-subject template estimation for  
625 unbiased longitudinal image analysis. *Neuroimage* **61**, 1402–1418 (2012).
- 626 56. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
- 627 57. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-Based Influences on  
628 Humans’ Choices and Striatal Prediction Errors. *Neuron* **69**, 1204–1215 (2011).

- 629 58. Pruim, R. H. R. *et al.* ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from  
630 fMRI data. *Neuroimage* **112**, 267–277 (2015).
- 631 59. Pruim, R. H. R., Mennes, M., Buitelaar, J. K. & Beckmann, C. F. Evaluation of ICA-AROMA and  
632 alternative strategies for motion artifact removal in resting state fMRI. *Neuroimage* **112**, 278–287  
633 (2015).
- 634 60. Hsu, C.-W., Chang, C.-C. & Lin, C.-J. *A Practical Guide to Support Vector Classification*. (2016).  
635 doi:10.1177/02632760022050997
- 636 61. Cortes, C. & Vapnik, V. Support-Vector Networks. *Mach. Learn.* **20**, 273–297 (1995).
- 637 62. Guggenmos, M., Schmack, K. & Sterzer, P. WeiRD - a fast and performant multivoxel pattern classifier.  
638 *2016 Int. Work. Pattern Recognit. Neuroimaging* 1–4 (2016). doi:10.1109/PRNI.2016.7552349
- 639 63. Guggenmos, M., Sterzer, P. & Cichy, R. M. Multivariate pattern analysis for MEG: a comparison of  
640 dissimilarity measures. *Neuroimage* **173**, 434–447 (2018).
- 641 64. Zhou, L., Wang, L., Liu, L., Ogunbona, P. O. & Shen, D. Support Vector Machines for neuroimage  
642 analysis: Interpretation from discrimination. in *Support Vector Machines Applications* 191–220  
643 (Springer, 2014).
- 644 65. Chang, C. & Lin, C. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst.*  
645 *Technol.* **2**, 27:1--27:27 (2011).
- 646 66. Luntz, A. & Brailovsky, V. On estimation of characters obtained in statistical procedure of recognition.  
647 *Tech. Kibern.* **3**, 6–12 (1969).
- 648 67. Brodersen, K. H., Ong, C. S., Stephan, K. E. & Buhmann, J. M. The balanced accuracy and its posterior  
649 distribution. *Proc. - Int. Conf. Pattern Recognit.* 3121–3124 (2010). doi:10.1109/ICPR.2010.764
- 650 68. Dwyer, D. B., Falkai, P. & Koutsouleris, N. Machine Learning Approaches for Clinical Psychology and  
651 Psychiatry. *Ssrn* 1–28 (2018). doi:10.1146/annurev-clinpsy-032816-045037
- 652 69. Haxby, J. V., Connolly, A. C. & Guntupalli, J. S. Decoding Neural Representational Spaces Using  
653 Multivariate Pattern Analysis. *Annu. Rev. Neurosci.* 435–456 (2014). doi:10.1146/annurev-neuro-  
654 062012-170325
- 655 70. Jamalabadi, H., Alizadeh, S., Schönauer, M., Leibold, C. & Gais, S. Classification based hypothesis  
656 testing in neuroscience: Below-chance level classification rates and overlooked statistical properties of  
657 linear parametric classifiers. *Hum. Brain Mapp.* **37**, 1842–1855 (2016).
- 658 71. Goldstein, R. Z. & Volkow, N. D. Dysfunction of the prefrontal cortex in addiction: neuroimaging  
659 findings and clinical implications. *Nat. Rev. Neurosci.* **12**, 652–669 (2012).
- 660 72. Abernathy, K., Chandler, L. J. & Woodward, J. J. Alcohol and the Prefrontal Cortex. *Int. Rev.*  
661 *Neurobiol.* **91**, 289–320 (2010).
- 662 73. Perry, J. L. *et al.* Prefrontal cortex and drug abuse vulnerability: Translation to prevention and treatment  
663 interventions. *Brain Res. Rev.* **65**, 124–149 (2011).
- 664 74. Ikemoto, S. & Panksepp, J. The role of nucleus accumbens dopamine in motivated behavior: A unifying  
665 interpretation with special reference to reward-seeking. *Brain Res. Rev.* **31**, 6–41 (1999).
- 666 75. Volkow, N. D., Wang, G.-J., Fowler, J. S. & Tomasi, D. Addiction Circuitry in the Human Brain. *Annu.*  
667 *Rev. Pharmacol. Toxicol.* **52**, 321–336 (2012).
- 668 76. Goldstein, R. Z. & Volkow, N. D. Drug addiction and its underlying neurobiological basis:

- 669 Neuroimaging evidence for the involvement of the frontal cortex. *Am. J. Psychiatry* **159**, 1642–1652  
670 (2002).
- 671 77. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **20**, 37–46 (1960).
- 672 78. Landis, R. J. & Koch, G. G. The Measurement of Observer Agreement for Categorical Data. *Biometrics*  
673 **33**, 159–174 (1977).
- 674 79. Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F. & Baker, C. I. Circular analysis in systems  
675 neuroscience: the dangers of double dipping. *Nat. Neurosci.* **12**, 535–40 (2009).
- 676 80. Cui, Y. *et al.* Identification of conversion from mild cognitive impairment to alzheimer’s disease using  
677 multivariate predictors. *PLoS One* **6**, 2–11 (2011).
- 678 81. Pettersson-Yeo, W. *et al.* An empirical comparison of different approaches for combining multimodal  
679 neuroimaging data with support vector machine. *Front. Neurosci.* **8**, (2014).
- 680 82. Ray, B. *et al.* Information content and analysis methods for multi-modal high-throughput biomedical  
681 data. *Sci. Rep.* **4**, 1–10 (2014).
- 682 83. Esteban, O. *et al.* FMRIprep: a robust preprocessing pipeline for functional MRI. *bioRxiv* (2018).  
683 doi:10.1101/306951
- 684 84. Agartz, I., Shoaf, S., Rawlings, R. R., Momenan, R. & Hommer, D. W. CSF monoamine metabolites  
685 and MRI brain volumes in alcohol dependence. *Psychiatry Res. - Neuroimaging* **122**, 21–35 (2003).
- 686 85. Hommer, D. W., Momenan, R., Kaiser, E. & Rawlings, R. R. Evidence for a gender-related effect of  
687 alcoholism on brain volumes. *Am. J. Psychiatry* **158**, 198–204 (2001).
- 688 86. Schweinsburg, B. C. *et al.* Effects of alcoholism and gender on brain metabolism. *Am. J. Psychiatry*  
689 **160**, 1180–1183 (2003).
- 690 87. Sullivan, E. V & Pfefferbaum, A. Neuropsychology and neuroimaging studies in alcohol-dependence 30  
691 ans de recherche en neuropsychologie et en imagerie cérébrale sur l’alcoolodépendance. **5**, 187–199  
692 (2013).
- 693 88. Fede, S. J., Grodin, E. N., Dean, S. F., Diazgranados, N. & Momenan, R. Resting state connectivity best  
694 predicts alcohol use severity in moderate to heavy alcohol users. *NeuroImage Clin.* **22**, 101782 (2019).
- 695 89. Demirakca, T. *et al.* Effects of Alcoholism and Continued Abstinence on Brain Volumes in Both  
696 Genders. *Alcohol. Clin. Exp. Res.* **35**, 1678–1685 (2011).
- 697 90. Cardenas, V. A., Studholme, C., Gazdzinski, S., Durazzo, T. C. & Meyerhoff, D. J. Deformation-based  
698 morphometry of brain changes in alcohol dependence and abstinence. *Neuroimage* **34**, 879–887 (2007).
- 699 91. Sullivan, E. V., Rosenbloom, M. J., Lim, K. O. & Pfefferbaum, A. Longitudinal changes in cognition,  
700 gait, and balance in abstinent and relapsed alcoholic men: Relationships to changes in brain structure.  
701 *Neuropsychology* **14**, 178–188 (2000).
- 702 92. Whelan, R. *et al.* Neuropsychosocial profiles of current and future adolescent alcohol misusers. *Nature*  
703 **512**, 185–189 (2014).
- 704 93. National Institute of Mental Health. Research Domain Criteria (RDoC). (2013).
- 705 94. Yücel, M. *et al.* A transdiagnostic dimensional approach towards a neuropsychological assessment for  
706 addiction: an international Delphi consensus study. *Addiction* **114**, 1095–1109 (2019).
- 707 95. Sayette, M. A. *et al.* The measurement of drug craving. *Addiction* **95**, 189–210 (2000).
- 708 96. Conigrave, K. M. *et al.* CDT, GGT, and AST as markers of alcohol use: The WHO/ISBRA

- 709 Collaborative Project. *Alcohol. Clin. Exp. Res.* **26**, 332–339 (2002).
- 710 97. Saunders, J. B. *et al.* Development of the Alcohol Use Disorders Identification Test (AUDIT). *Addiction*  
711 791–804 (1993).
- 712 98. Schmidt, L., Gastpar, M., Falkai, P. & Gäbel, W. Evidenzbasierte Suchtmedizin. in *Substanzbezogene*  
713 *Störungen* (Deutscher Ärzteverlag, 2006).
- 714 99. Skinner, H. & Horn, J. *Alcohol Dependence Scale (ADS): User's Guide.* (1984).
- 715 100. Mann, K. & Ackermann, K. Die OCDS-G: Psychometrische Kennwerte der deutschen Version der  
716 Obsessive Compulsive Drinking Scale [The OCDS-G: Psychometric Characteristics of the German  
717 version of the Obsessive Compulsive Drinking Scale]. *Sucht* **46**, 90–100 (2000).
- 718 101. Meule, A., Vögele, C. & Kübler, A. Psychometrische evaluation der Deutschen Barratt Impulsiveness  
719 scale - Kurzversion (BIS-15). *Diagnostica* **57**, 126–133 (2011).
- 720 102. *Army Individual Test Battery. Manual of directions and scoring. War Department, Adjutant General's*  
721 *Office.* (1944).
- 722 103. Wechsler, D. *WAIS-III, Wechsler Adult Intelligence Scale: Administration and Scoring Manual.* (1997).
- 723 104. Lehrl, S. *Mehrfachwahl-Wortschatz-Intelligenztest MWT-B.* (Spitta, 2005).
- 724 105. Morris, J. C. *et al.* The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) - Part I.  
725 Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* **39**, 1159–65 (1989).
- 726

727 **Figure Legends**

728 **Figure 1. Multimodal classification scheme.** Depicted is one exemplary split into training  
729 data and test data. Using a nested optimization loop, three modality-specific factors are  
730 optimized on the training data: classifier types (SVM, WeiRD), parameters (cost parameter  $C$   
731 for SVM) and weights  $w_i$ . The trained and optimized model is then applied to the test data  
732 and continuous decision values  $d_i$  are computed for each modality-specific classifier. The  
733 final diagnostic classification is based on a weighted sum of decision values, where weights  
734 correspond to those estimated during training.

735 **Figure 2. Unimodal classification.** Feature importances of **(A)** structural and **(B)** functional  
736 neuroimaging modalities. Depicted are 2-d projections ('glass brains') of feature importances  
737 along the x- and z-axis. Feature importances represent SVM weights (grey-matter density,  
738 reward response) or WeiRD votes (cerebrospinal fluid, cortical thickness, resting state)  
739 depending on which classifiers was superior for a given modality. **(C)** Inter-modality  
740 reliability matrix based on Cohen's Kappa describing the diagnostic agreement between  
741 modalities.

742 **Figure 3. Multimodal classification.** Balanced accuracy for classification schemes based on  
743 different classifier configurations (SVM, WeiRD or optimized between SVM and WeiRD)  
744 and with uniform (i.e., all weights set to 1) or optimized weighting of modalities. Optimizing  
745 both classifiers and weighting yielded the best performance (highlighted in green). Red lines  
746 indicate the balanced accuracy when ensemble prediction was based on discrete “control” and  
747 “patient” judgements instead of continuous decision values. Error bars represent the 95%  
748 posterior probability interval<sup>67</sup>.

749

750 **Tables**751 **Table 1.** Sample characteristics for alcohol-dependent (AD) and healthy control (HC) subjects.

	AD (N = 119)			HC (N = 97)			<i>t</i> or $\chi^2$	df	p
	Mean	SD	%	Mean	SD	%			
<b>Gender [female]</b>			15.1			16.5	0.008	N=216	.93
<b>Age [years]</b>	45.0	10.7		43.6	10.8		0.9	214	.38
<b>Education [years]</b>	10.5	0.1		11.2	0.2		-3.4	207	<.001
<b>SES</b>	-0.4	0.2		0.7	0.3		-3.6	170	<.001
<b>Smokers</b>			76.5			67.0	1.9	N=216	.16
<b>ADS score</b>	14.8	6.9		2.0	3.0		17.0	213	<.001
<b>AD duration [years]</b>	11.7	9.9						N=110	
<b>Amount life [kg]</b>	1805	1121		286	811		11.1	214	<.001
<b>Amount past year [kg]</b>	178	13		11	1		12.0	214	<.001
<b>OCDS total score</b>	11.9	8.5		2.8	2.8		10.1	207	<.001
<b>BIS-15 total score</b>	31.6	6.5		29.1	5.5		2.9	205	.004
<b>TMT (percentile)</b>	36.1	25.1		44.8	25.1		2.5	209	.014
<b>DSST</b>	64.3	15.1		73.5	16.6		4.2	211	<.001
<b>DSB</b>	6.5	1.9		7.4	2.0		3.4	214	.001
<b>MWT</b>	104.7	9.4		104.5	8.9		-0.2	209	.82
<b>Wordlist</b>	90.8	16.1		90.9	14.1		-0.0	209	.97

752 Socioeconomic status (SES): sum of z-transformed self-ratings of social status, household income and inverse personal debt  
 753 scores <sup>98</sup>; Alcohol Dependence Scale (ADS): degree/level of AD <sup>99</sup>; Amount life: lifetime alcohol consumption in kilograms  
 754 based on the CAPI-CIDI (Wittchen and Pfister, 2007; Jacobi et al., 2013); Amount past year: alcohol consumption during  
 755 the past year in kilograms based on the CAPI-CIDI (Wittchen and Pfister, 2007; Jacobi et al., 2013); Obsessive Compulsive  
 756 Drinking Scale (OCDS): Current craving for alcohol <sup>100</sup>; Barratt Impulsiveness scale (BIS-15): impulsivity <sup>101</sup>; Trail making  
 757 test (TMT; percentile): visual attention and task switching (Army Individual Test Battery, 1944); Digit symbol substitution  
 758 test (DSST): processing speed <sup>102</sup>. Digit span backwards (DSB): working memory span <sup>103</sup>. Multiple-choice vocabulary  
 759 intelligence test (Mehrfachwahl-Wortschatz-Intelligenztest, MWT): crystallized / verbal intelligence <sup>104</sup>; Wordlist (savings):  
 760 wordlist memory test <sup>105</sup>.

761  
762  
763

**Table 2. Overview of modalities.** Columns represent acquisition time, shortcuts for each modality used throughout the article, the number of control ( $N_{\text{Ctr}}$ ) and patients ( $N_{\text{Pat}}$ ) available for each modality, and the numbers of features per modality (No. features).

	<b>Time</b>	<b>Modality</b>	<b>Short</b>	$N_{\text{Ctr}}$	$N_{\text{Pat}}$	<b>No. features</b>
<b>sMRI</b>	4:26	Grey-matter density	GMD	97	119	110
		Cerebrospinal fluid	CSF	97	119	11
		Cortical thickness	CTH	96	119	358
<b>Task-based fMRI</b>	22:10	Reward response	RWR	74	80	1461
<b>Resting-state fMRI</b>	6:02	Nucleus accumbens connectivity	NAC	84	93	110

764

765

766  
767  
768

**Table 3. Unimodal classification.** Classification performance of the SVM and WeiRD classifiers for the five modalities under consideration. Abbreviations: Acc. = Balanced accuracy; Sens. = sensitivity; Spec. = specificity; NAcc = Nucleus Accumbens.

	<b>SVM</b>				<b>WeiRD</b>			
	<b>Acc.</b>	<b>Sens.</b>	<b>Spec.</b>	<b>p</b>	<b>Acc.</b>	<b>Sens.</b>	<b>Spec.</b>	<b>p</b>
<b>Grey-matter density</b>	76.6	79.0	74.2	<.001	71.3	66.4	76.3	<.001
<b>Cerebrospinal fluid</b>	58.6	51.3	66.0	.003	65.0	58.8	71.1	<.001
<b>Cortical thickness</b>	54.9	58.8	51.0	.037	65.6	69.7	61.5	<.001
<b>Reward response</b>	60.2	47.5	73.0	.002	59.3	55.0	63.5	.005
<b>NAcc connectivity</b>	54.8	54.8	54.8	.050	55.0	50.5	59.5	.044

769

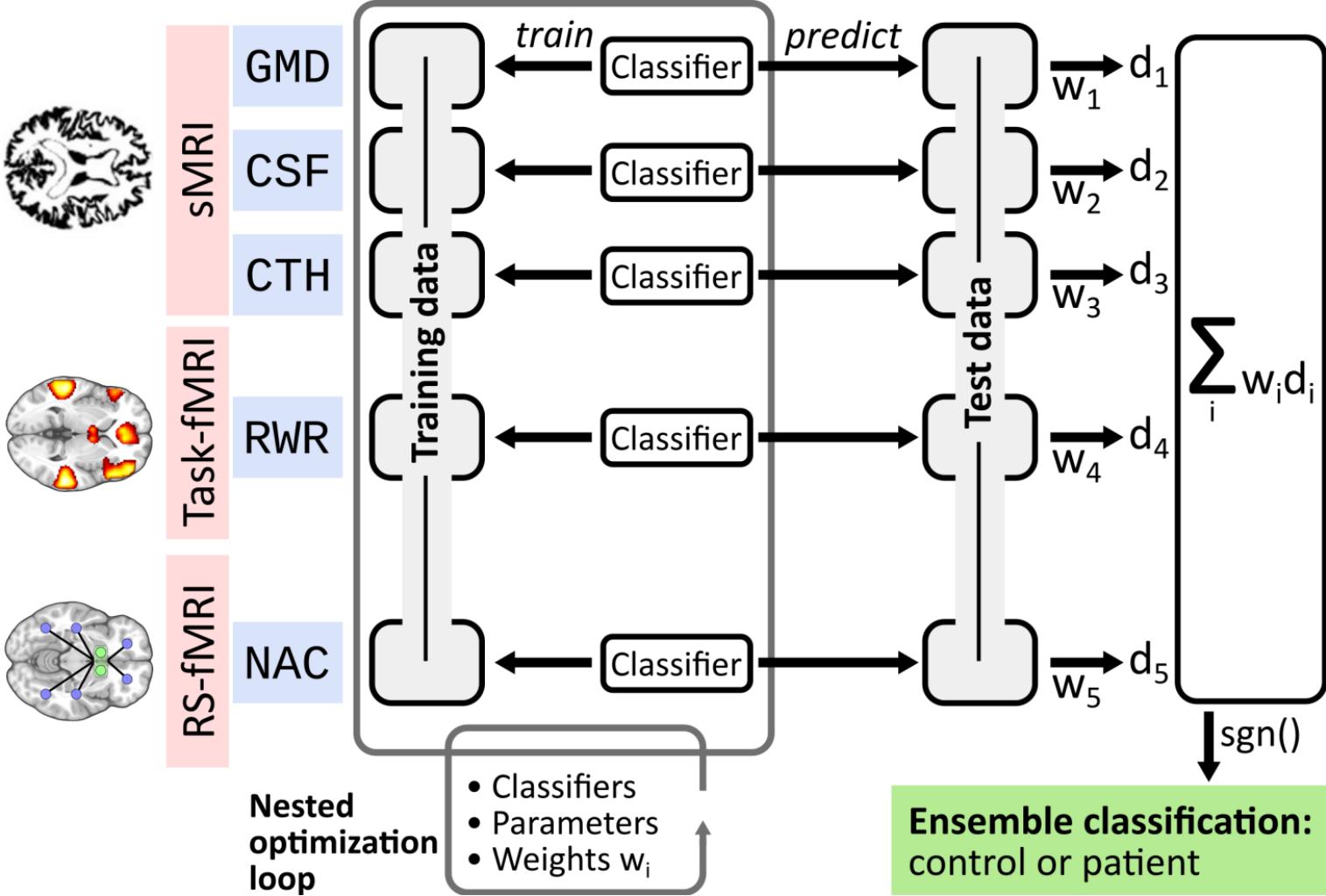
770

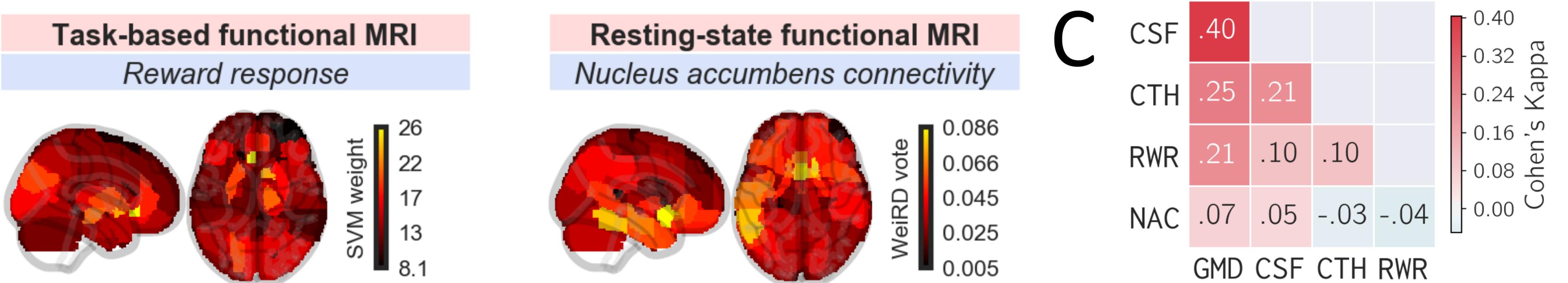
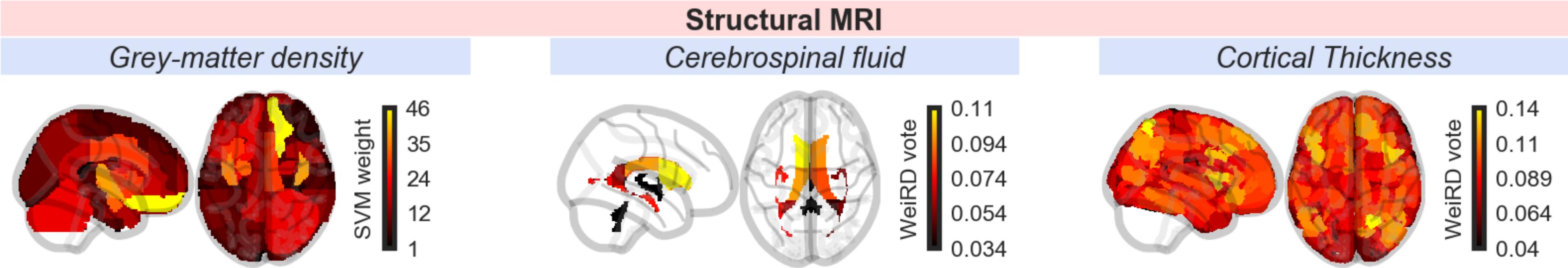
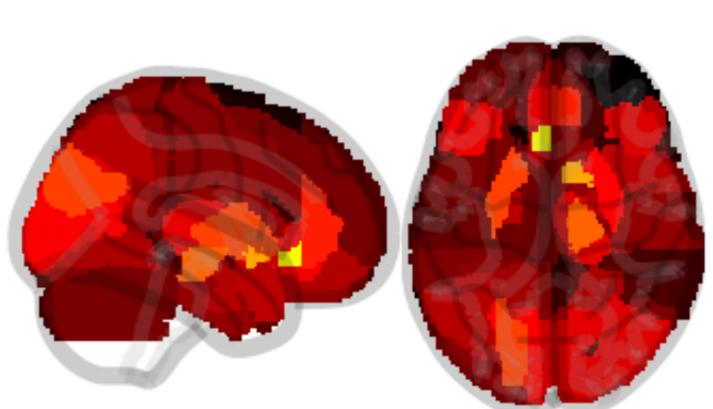
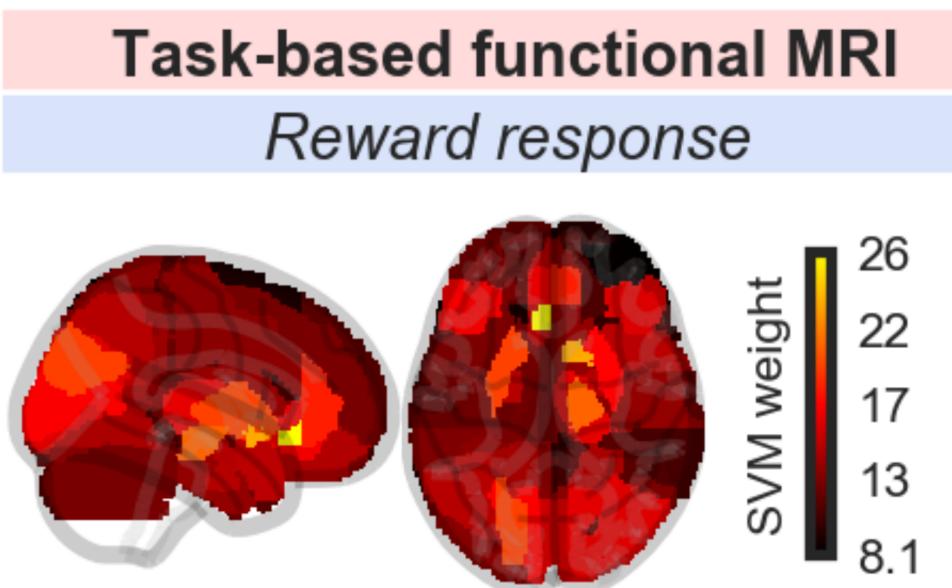
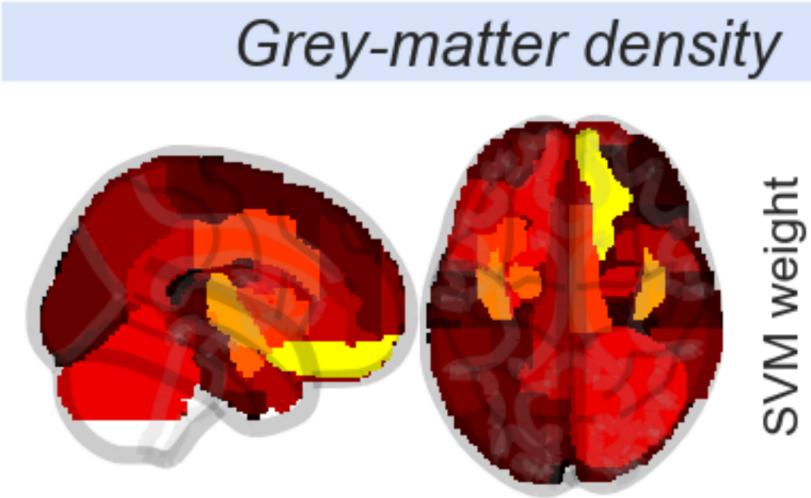
771  
772  
773

**Table 4. Impact of excluding modalities.** Performance of multimodal classification when leaving out each modality once ( $\Delta$  = change with respect to the full model). Abbreviations: Acc. = Balanced accuracy; Sens. = sensitivity; Spec. = specificity; NAcc = Nucleus Accumbens.

	<b>Acc. (<math>\Delta</math>)</b>	<b>Sens. (<math>\Delta</math>)</b>	<b>Spec. (<math>\Delta</math>)</b>
<b>Grey-matter density</b>	69.7 (-9.6)	59.2 (-18.1)	80.2 (-1.0)
<b>Cerebrospinal fluid</b>	75.3 (-4.0)	76.7 (-0.6)	74.0 (-7.3)
<b>Cortical thickness</b>	78.1 (-1.2)	75.8 (-1.5)	80.4 (-0.8)
<b>Reward response</b>	77.0 (-2.3)	75.6 (-1.7)	78.4 (-2.9)
<b>NAcc connectivity</b>	77.3 (-2.0)	77.3 (0.0)	77.3 (-3.9)

774



**A****C**

	GMD	CSF	CTH	RWR
CSF	.40			
CTH	.25	.21		
RWR	.21	.10	.10	
NAC	.07	.05	-.03	-.04

Cohen's Kappa

**Classifier:** WeiRD  
**Weighting:** uniform

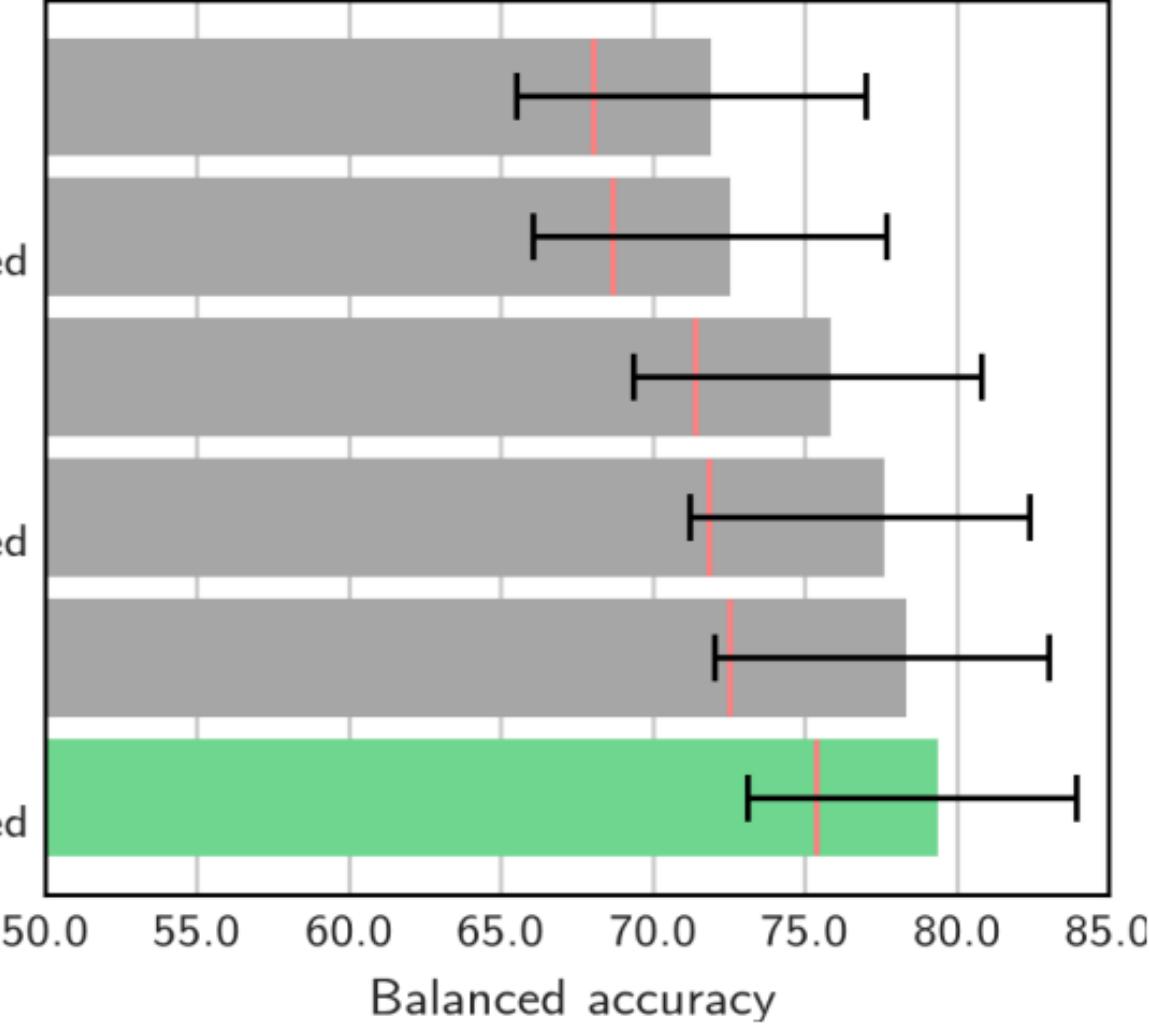
**Classifier:** WeiRD  
**Weighting:** optimized

**Classifier:** SVM  
**Weighting:** uniform

**Classifier:** SVM  
**Weighting:** optimized

**Classifier:** optimized  
**Weighting:** uniform

**Classifier:** optimized  
**Weighting:** optimized



## **Supplementary material to: A multimodal neuroimaging classifier for alcohol dependence**

Matthias Guggenmos<sup>1</sup>, Katharina Schmack<sup>1</sup>, Ilya M. Veer<sup>1</sup>, Tristram Lett<sup>1</sup>, Maria Sekutowicz<sup>1</sup>, Miriam Sebold<sup>1</sup>, Maria Garbusow<sup>1</sup>, Christian Sommer<sup>2</sup>, Hans-Ulrich Wittchen<sup>4,5</sup>, Ulrich S. Zimmermann<sup>2</sup>, Michael N Smolka<sup>2,3</sup>, Henrik Walter<sup>1</sup>, Andreas Heinz<sup>1</sup>, Philipp Sterzer<sup>1</sup>

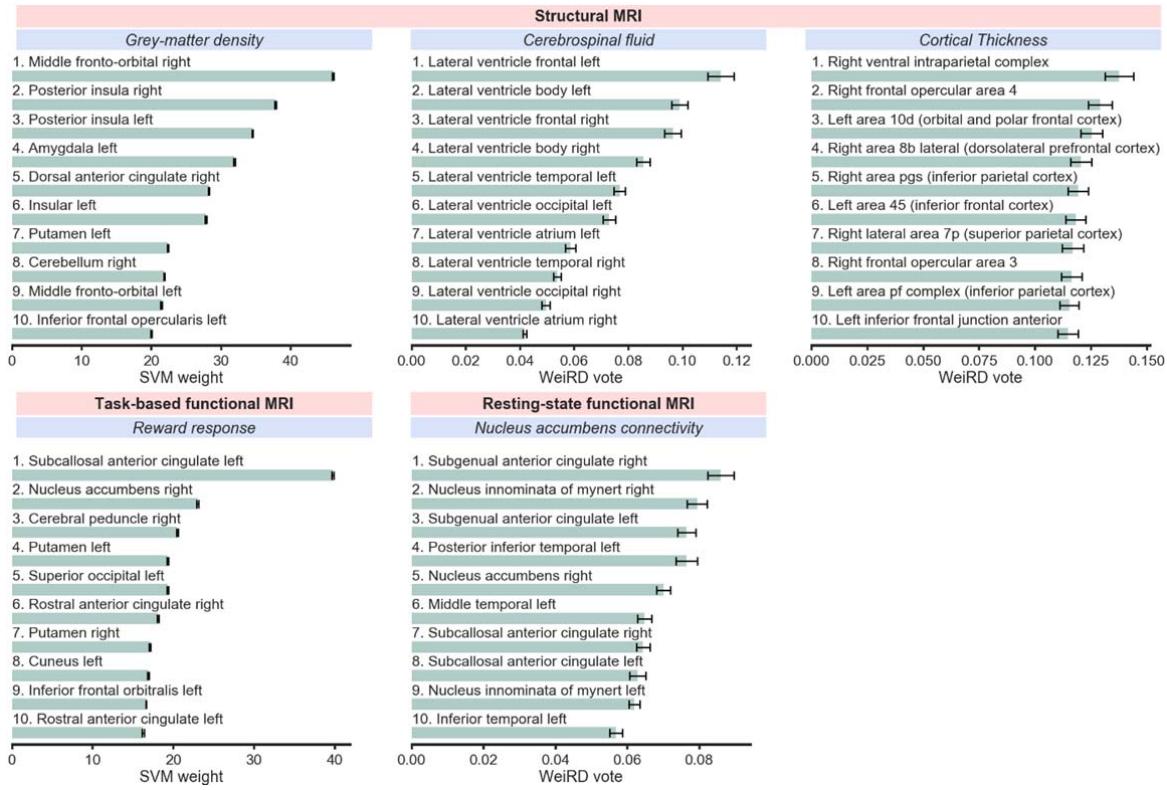
<sup>1</sup> Department of Psychiatry and Psychotherapy, Charité Universitätsmedizin, Germany

<sup>2</sup> Department of Psychiatry and Psychotherapy, Technische Universität Dresden, Germany

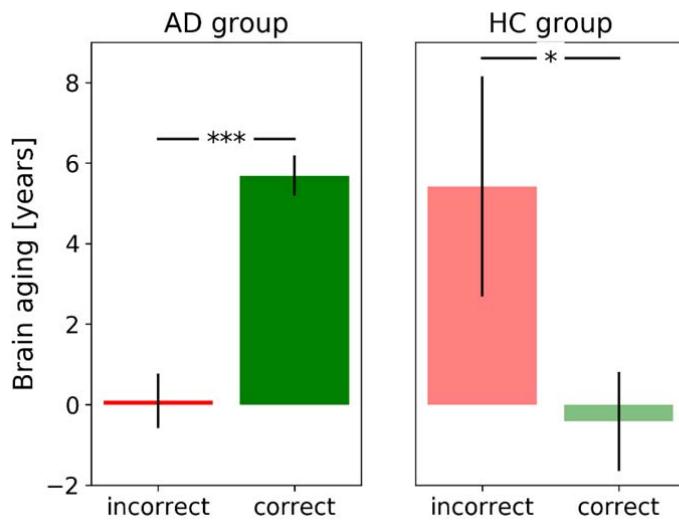
<sup>3</sup> Neuroimaging Center, Technische Universität Dresden, Dresden, Germany

<sup>4</sup> Institute of Clinical Psychology and Psychotherapy, Technische Universität Dresden, Germany

<sup>5</sup> Department of Psychiatry and Psychotherapy, Ludwig Maximilians Universität Munich, Munich, Germany



**Supplementary Figure S1. Feature importances.** Depicted are for each modality the 10 most important brain regions for unimodal classification. Feature importance is characterized by SVM weights in case of SVM classification (grey-matter density, reward responses) and by WeiRD votes in case of WeiRD classification (cerebrospinal fluid, cortical thickness, resting state connectivity). Error bars represent the 95% confidence interval across cross-validation folds.



**Supplementary Figure S2. Relationship between brain aging and classifier predictions.** Brain aging quantifies the difference between chronological age and biological age estimated on the basis of grey-matter volume. Brain aging estimates are derived from a previous study based on the same sample [Guggenmos et al., 2017]. Quantitative neurobiological evidence for accelerated brain aging in alcohol dependence. *Transl. Psychiatry* 7. doi: 10.1038/s41398-017-0037-y]. Brain aging in correctly predicted patients ( $5.7 \pm 0.5$  years) was significantly more pronounced than in patients incorrectly predicted as controls ( $0.1 \pm 0.7$  years) (paired t-test:  $t_{117} = -5.6$ ,  $p < 0.001$ ). In controls, this pattern of results was reversed: brain aging was higher in incorrectly ( $5.4 \pm 2.7$  years) compared to incorrectly ( $-0.42 \pm 1.2$  years) classified controls ( $t_{94} = 2.0$ ,  $p=0.045$ ). This provides suggestive evidence that predictions of the multimodal classifier were largely based on neurotoxic consequences of alcohol dependence.