

OPEN

A multimodal neuroimaging classifier for alcohol dependence

Matthias Guggenmos^{1*}, Katharina Schmack¹, Ilya M. Veer¹, Tristram Lett¹, Maria Sekutowicz¹, Miriam Sebold¹, Maria Garbusow¹, Christian Sommer², Hans-Ulrich Wittchen^{4,5}, Ulrich S. Zimmermann², Michael N. Smolka^{2,3}, Henrik Walter¹, Andreas Heinz¹ & Philipp Sterzer¹

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

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

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

Here, we reasoned that alcohol dependence is a well-suited *psychiatric* disorder for automated diagnostic classification based on multimodal MRI and an ideal test case and benchmark for methodological developments. This is first and foremost because neurobiological correlates of alcohol dependence have been established at several levels of description, including grey-matter loss^{15–18}, increased ventricular size/cerebrospinal fluid concentration^{19–21} and aberrant neural reward responses^{22–25}. The diversity of effects in combination with a high degree of

¹Department of Psychiatry and Psychotherapy, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany. ²Department of Psychiatry and Psychotherapy, Technische Universität Dresden, Dresden, Germany. ³Neuroimaging Center, Technische Universität Dresden, Dresden, Germany. ⁴Institute of Clinical Psychology and Psychotherapy, Technische Universität Dresden, Dresden, Germany. ⁵Department of Psychiatry and Psychotherapy, Ludwig Maximilians Universität Munich, Munich, Germany. *email: matthias.guggenmos@charite.de

	AD (N = 119)			HC (N = 97)			t or χ^2	df	p
	Mean	SD	%	Mean	SD	%			
Gender [female]			15.1			16.5	0.008	N = 216	0.93
Age [years]	45.0	10.7		43.6	10.8		0.9	214	0.38
Education [years]	10.5	0.1		11.2	0.2		-3.4	207	<0.001
SES	-0.4	0.2		0.7	0.3		-3.6	170	<0.001
Smokers			76.5			67.0	1.9	N = 216	0.16
ADS score	14.8	6.9		2.0	3.0		17.0	213	<0.001
AD duration [years]	11.7	9.9						N = 110	
Amount life [kg]	1805	1121		286	811		11.1	214	<0.001
Amount past year [kg]	178	13		11	1		12.0	214	<0.001
OCDS total score	11.9	8.5		2.8	2.8		10.1	207	<0.001
BIS-15 total score	31.6	6.5		29.1	5.5		2.9	205	0.004
TMT (percentile)	36.1	25.1		44.8	25.1		2.5	209	0.014
DSST	64.3	15.1		73.5	16.6		4.2	211	<0.001
DSB	6.5	1.9		7.4	2.0		3.4	214	0.001
MWT	104.7	9.4		104.5	8.9		-0.2	209	0.82
Wordlist	90.8	16.1		90.9	14.1		-0.0	209	0.97

Table 1. Sample characteristics for alcohol-dependent (AD) and healthy control (HC) subjects. Socioeconomic status (SES): sum of z-transformed self-ratings of social status, household income and inverse personal debt scores⁹⁸; Alcohol Dependence Scale (ADS): degree/level of AD⁹⁹; Amount life: lifetime alcohol consumption in kilograms based on the CAPI-CIDI (Wittchen and Pfister, 2007; Jacobi *et al.*, 2013); Amount past year: alcohol consumption during the past year in kilograms based on the CAPI-CIDI (Wittchen and Pfister, 2007; Jacobi *et al.*, 2013); Obsessive Compulsive Drinking Scale (OCDS): Current craving for alcohol¹⁰⁰; Barratt Impulsiveness scale (BIS-15): impulsivity¹⁰¹; Trail making test (TMT; percentile): visual attention and task switching¹⁰²; Digit symbol substitution test (DSST): processing speed¹⁰³. Digit span backwards (DSB): working memory span¹⁰³. Multiple-choice vocabulary intelligence test (Mehrfachwahl-Wortschatz-Intelligenztest, MWT): crystallized / verbal intelligence¹⁰⁴; Wordlist (savings): wordlist memory test¹⁰⁵.

replicability could make the combination of different MRI modalities particularly powerful in the case of alcohol dependence. In addition, alcohol dependence, relative to other psychiatric disorders, is a reliable diagnosis^{26–28} and thus a paradigmatic case to gauge the true predictive potential of an MRI-based classifier for psychiatric diagnosis.

Our investigation was based on a sample of 119 alcohol-dependent patients and 97 controls who underwent an extensive neuroimaging battery including structural^{16,29}, functional^{30–32} and resting-state³³ MRI as part of the LeAD study (www.lead-studie.de; clinical trial number: NCT01679145). Our guiding rationale for the construction of a multimodal diagnostic classifier was clinical utility and practicability, and thus to leverage on neuroimaging modalities that are effective and simple to acquire (i.e., relatively short acquisition times and standard scanning sequences) and to analyse (e.g. no computational cognitive modelling or other approaches rich on assumptions). Selected modalities comprised grey-matter density, cerebral spinal fluid and cortical thickness based on structural MRI as well as basic reward responses and resting-state connectivity based on functional MRI.

Our goal was to address key methodological challenges of multimodal MRI, such as the heterogeneity of different modalities in terms of statistical properties and the number of predictors, missing data in one or more modalities, and the relative weighting of each modality. In the present investigation we thus developed a novel classification framework that (1) combined modality-specific predictions in an ensemble vote, (2) allowed for emphasizing or de-emphasizing individual modalities through weighting, (3) considered fine-grained information from modality-specific classifiers instead of binary labels, and (4) was robust to missing data in individual modalities. We hypothesized that the combination of MRI modalities for diagnostic classification of alcohol dependence would outperform any individual unimodal classification approach.

Method

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

We used the computer-assisted interview version Composite International Diagnostic Interview (CAPI-CIDI^{36,37}) to verify diagnosis criteria of AD in the patient group and to exclude the possibility of AD in control subjects. For inclusion, individuals with AD had to meet criteria for AD for at least three years and had to undergo an inpatient detoxification phase (average duration \pm SEM: 22.8 \pm 1 days). Exclusion criteria for all subjects were left-handedness (Edinburgh handedness index below 50³⁸), contraindications for MRI, and a history of or current neurological (including Korsakoff syndrome) or mental disorders (excluding nicotine dependence in both groups and alcohol abuse in individuals with AD, but including abuse of other drugs). Mental disorders were

assessed according to DSM-IV axis one as verified by the CAPI-CIDI. It was ensured that all subjects were free of psychotropic medication (including detoxification treatment) known to interact with the central nervous system for at least four half-lives. Current non-alcohol drug abuse was confirmed by means of a dedicated urine test.

Note that several sampling characteristics were influenced by the fact that predictors of relapse, while not part of the present work, were another research focus within the LeAD study³⁰: (1) to have sufficient power for the comparison between relapsers and abstainers, the AD group was oversampled relative to the HC group; (2) to be able to study relapse behaviour, detoxification at baseline was an inclusion criterion; (3) to prevent the possibility that handedness would be unequally distributed across relapsers and abstainers, only right handers were included. Finally note that matching for smoking implied a higher proportion of smokers in the HC sample than expected from a representative sample (67% versus a representative prevalence of 28% in Germany³⁹).

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

Overview of neuroimaging modalities. To maximize the clinical utility of a multimodal diagnostic classifier, we constructed a neuroimaging battery of five modalities that was (1) simple to acquire and to analyse, and (2) based on established effects either in previous literature or on own works within the LeAD program. From structural MRI measurements we derived three modalities: (1) grey-matter density, as damaging effects to grey-matter integrity in chronic alcoholics have been replicated numerous times in the literature^{15,17,18} including our own work in which we also demonstrated high diagnostic discriminability¹⁶; (2) cerebrospinal fluid, as increased ventricular size (or increased cerebrospinal fluid, respectively) is arguably one of the most salient characteristics of alcoholic brains^{19–21}; and (3) cortical thickness, an additional specific marker of grey-matter integrity obtained through surface-based analysis, which has been successfully used in more recent studies to characterize structural damage in alcohol dependence^{40–42}.

From task-based functional MRI we obtained (4) functional activation patterns representing a basic reward response (outcome versus no outcome), motivated by the fact that aberrant functional reward responses have been consistently found for alcohol dependence^{22–25}. Finally, from resting-state functional MRI, we derived (5) nucleus accumbens whole-brain connectivity maps, as one of the key research goals of the LeAD program was testing a hypothesis about disturbed striato-frontal connectivity underlying the development of problematic drinking behaviours (this hypothesis has recently been confirmed by authors of this article³³). Note that all data were linearly corrected for variance of no interest related to demographic variables age, gender and site.

MRI data acquisition. Magnetic resonance imaging (MRI) was performed on a 3-Tesla Siemens Trio (Erlangen, Germany) scanner with a 12-channel head-coil to obtain (1) structural, (2) functional task-based and (3) functional resting-state MRI data. Structural T₁-weighted MRI scans were acquired using a magnetization-prepared rapid gradient echo sequence (repetition time: 1900 ms; echo time: 5.25 ms; flip angle: 9°; field of view: 256 × 256 mm²; voxel size: 1 mm isotropic; 192 sagittal slices). Functional (task-based or resting-state) T₂*-weighted MRI scans were acquired using a gradient echo planar imaging sequence (repetition time: 2410 ms; echo time: 25 ms; flip angle: 80°; field of view: 192 × 192 mm²; voxel size: 3 × 3 × 2 mm³) comprising 42 slices approximately –25° to the bicommissural plane. Volume-to-volume movement of more than 3 mm translation and/or 2 degrees rotation led to exclusion (9 HC and 12 AD subjects were excluded due to these criteria). All imaging data were screened for corrupted data or serious acquisition artefacts.

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

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

To obtain estimates of CTH, cortical reconstruction was performed on T₁-weighted images using the FreeSurfer morphometric analysis suite (<http://surfer.nmr.mgh.harvard.edu/>). The technical details of these procedures are described in prior publications^{44–55}. CTH estimates were obtained by calculating the closest distance from the grey/white boundary to the grey/CSF boundary at each vertex on the tessellated surface⁴⁵. Average cortical thickness estimates were obtained for 358 brain regions based on the Glasser anatomical atlas⁵⁶.

Task-based functional MRI. Preprocessing was performed using SPM8 (www.fil.ion.ucl.ac.uk/spm) and included slice time correction, realignment to the first image, coregistration with the structural image, spatial normalization into MNI space and smoothing (8 mm Gaussian kernel). Details of the paradigm and the statistical first-level model have been provided elsewhere³⁰. In brief, the paradigm involved a probabilistic value-based decision making task⁵⁷ in which participants could receive a fixed monetary reward in each of 201 trials. While the goal of the paradigm is to distinguish model-based from model-free learning signals by means of a computational

model, following our rationale of constructing a neuroimaging battery with only the most fundamental modalities, we here used the basic contrast of reward (+20 Euro cents) versus no reward.

Within the statistical first-level model³⁰, we used the two onset regressors for *reward* and *no-reward* conditions during the outcome phase and computed directional contrasts (*reward* > *no reward*). To reduce the extremely high dimensionality of individual fMRI whole-brain contrast maps (>200,000 valid voxels), we downsampled the data to 10 mm isotropic voxels resulting in 1461 valid voxels (features) for each participant. Exploratory classification analyses with smaller voxels sizes (2–8 mm in steps of 2 mm) showed that no relevant information was lost at 10 mm.

Resting-state functional MRI. Preprocessing of resting-state data was performed using FSL (<https://www.fmrib.ox.ac.uk/fsl>) and included motion correction, slice timing correction, non-brain removal, 6 mm FWHM spatial smoothing. Data were denoised using independent component analysis (ICA-AROMA^{58,59}), high-pass-filtered at 0.008 Hz and normalized to MNI space. To estimate whole-brain accumbens connectivity maps, binary seed masks of the left and right nucleus accumbens were defined using the Harvard-Oxford Subcortical Probability Atlas (50% probability threshold). The first Eigen time series was extracted from the preprocessed resting-state data, separately for the left and right nucleus accumbens. Finally, these time series were regressed against every other voxel's time series using deep white matter and CSF as nuisance variables. To reduce dimensionality, we computed regional averages for the 110 anatomical regions of the JHU brain atlas⁴³ and averaged across the left and right nucleus accumbens. Note that while participants were excluded when falling asleep during scanning, part of the interindividual variance in resting state connectivity may be explained by unknown interindividual differences in drowsiness/vigilance, which were not assessed in an objective manner in the current study. Nevertheless, only a small number of participants (7 AUD and 6 HC) reported brief moments of drowsiness or loss of vigilance during the exit interview.

Unimodal diagnostic classification. In a first step, the diagnostic accuracy of each of the five modalities was assessed. Before submitting the modalities to classification, each modality was soft-normalized by removing the median and scaling the data according to a quantile range (1% to 99%)⁶⁰. We considered two different classifier types: support vector machine (SVM⁶¹) and weighted robust distance (WeiRD^{16,62,63}). SVM is arguably the most popular classifier in the context of neuroimaging⁶⁴, as it robustly handles data with high dimensionality but few samples per class. We used the implementation provided by libsvm⁶⁵ with a radial basis function kernel. 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 prototypes in a statistically weighted feature space. A key advantage of WeiRD is the fact that it is parameter-free and hence does not require expensive optimization schemes. In previous work we have shown that WeiRD, despite its simplicity, performs surprisingly well across a range of neuroimaging classification problems^{16,62,63}.

Both classifiers operated in a leave-one-out (LOO) cross-validation procedure, such that in each cross-validation fold, the classifiers were trained on the data of all but one participant and tested on the left-out participant. Each participant was left out and predicted exactly once. LOO was chosen over other cross-validation schemes, because it is approximately unbiased⁶⁶, deterministic and computationally inexpensive. For SVM specifically, we additionally performed nested cross-validation and grid search to optimize the cost parameter C (range of C : 2^x , $x = -5:10$). We used the balanced accuracy as a scoring metric and computed p-values of the null hypothesis that the balanced accuracy is equal to chance level⁶⁷.

Note that the number of features outnumbered the number of samples. This is the standard scenario in multivariate neuroimaging analyses, which have been widely used with success to fit machine learning models within and across individuals^{68,69}, most likely due to intercorrelations between features⁷⁰.

In addition, we assessed the importance of each region for classification. For clarity and brevity, we show importance scores only for the superior classifier (SVM or WeiRD) of a modality. Where SVM was superior, we derived importance scores from the hyperplane-defining weight vector in feature space. Although the feature space is not the input space in the case of the radial-basis function kernel used here, the weights in feature space nevertheless provided a useful estimate of feature importance for descriptive purposes. Where WeiRD was superior, we used WeiRD votes as a measure of feature importance, as described previously¹⁶.

Multimodal diagnostic classification. The aim of multimodal neuroimaging classification is to improve overall performance by leveraging on the combined information of more than one neuroimaging modality. Here we combined all five modalities introduced above (GMD, CSE, CTH, RWR, NAC) in an integrated classification scheme. The key design principle for the multimodal classification scheme was that a separate classifier was trained on each modality, which made it possible to select the optimal classifier for each modality and to fit it to the specific statistical properties of a modality.

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

Training. During training, a separate classifier was trained on each modality and three factors were optimized in a nested cross-validation procedure: the classifier type (SVM or WeiRD); the parameters of a classifier if applicable (cost parameter C for SVM); and modality-specific weights. In detail, for each modality separately it was first assessed whether SVM (with second-level-nested cross-validation for the optimization of C) or WeiRD were more accurate on the training data set. If SVM was superior, the optimal cost parameter was then estimated anew on the entire training data. After determining the optimal classifier for each modality, a new

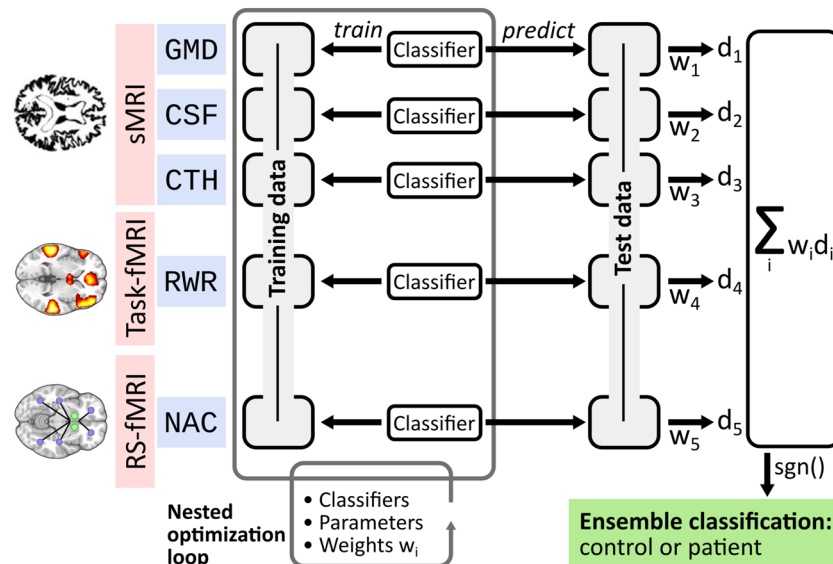


Figure 1. Multimodal classification scheme. Depicted is one exemplary split into training data and test data. Using a nested optimization loop, three modality-specific factors are optimized on the training data: classifier types (SVM, WeiRD), parameters (cost parameter C for SVM) and weights w_i . The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

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

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

nested cross-validation procedure was performed for all modalities combined to estimate weighting factors for each modality. To limit computational complexity and increase robustness, each of the five modalities could be weighted only with a factor of either 1 or 2.

Prediction. After training, the classification scheme was applied to the yet unseen test subject. Each modality-specific classifier computed a signed continuous decision value for the respective modality, where negative and positive values represented control and patient predictions and the absolute value the certainty of a classifier. The overall ensemble prediction was based on the sign of a weighted sum of modality-specific decision values with weighting factors determined during training.

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

Finally, note that a key advantage of our classification scheme was that missing data were of no concern: if a subject had missing data for a modality (see Table 2 for the number of valid samples in each modality), this modality was simply omitted from the weighted sum of modality-specific decision values for the given subject.

Results

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

We found that each modality discriminated significantly between patients and controls (Table 3), although the accuracy varied strongly across modalities. For both SVM and WeiRD, grey-matter density was the

	SVM				WeiRD			
	Acc.	Sens.	Spec.	P	Acc.	Sens.	Spec.	P
Grey-matter density	76.6	79.0	74.2	<0.001	71.3	66.4	76.3	<0.001
Cerebrospinal fluid	58.6	51.3	66.0	0.003	65.0	58.8	71.1	<0.001
Cortical thickness	54.9	58.8	51.0	0.037	65.6	69.7	61.5	<0.001
Reward response	60.2	47.5	73.0	0.002	59.3	55.0	63.5	0.005
NAcc connectivity	54.8	54.8	54.8	0.050	55.0	50.5	59.5	0.044

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.

best-performing modality with balanced accuracies of 76.6% and 71.3%, respectively. All other modalities ranged between 55 and 66% accuracy. The balance between specificity and sensitivity was heterogeneous across modalities. For instance, while cortical thickness was more sensitive than specific, cerebrospinal fluid and reward responses showed an opposite pattern.

To investigate which brain regions contributed to classification, we inspected feature importances for each modality (Fig. 2A,B and Supplementary Fig. S1). Across modalities, we found that anterior cingulate and inferior frontal brain regions discriminated best between patients and controls. This included the cerebrospinal fluid, for which the (left) frontal ventricle was most discriminative. These results are broadly in line with the hypothesis of prefrontal cortex dysfunction in addiction^{71–73}. In addition, for functional reward responses the nucleus accumbens was the second most discriminative brain region (after subcallosal anterior cingulate cortex), a key region of generic reward pathways⁷⁴ and of bottom-up theories of addiction such as the incentive-sensitization theory²². Although not the focus of this study, the brain regions that discriminate best between patients and controls map well on established neural correlates of alcohol dependence^{75,76} and provide validation to the approach taken in this study.

Our initial unimodal analysis of feature importances showed that informative features were primarily located in prefrontal and cingulate brain regions (for structural and connectivity measures) as well as in the nucleus accumbens (for reward responses).

How well do different modalities agree with respect to their diagnostic predictions? To find out, we computed inter-modality reliability scores based on Cohen's Kappa⁷⁷, which measures the agreement of two "raters" over and above the agreement expected by chance. For each modality we used the classifier (SVM or WeiRD) that performed better for a given modality. As shown in Fig. 2C, the agreement was generally highest between modalities based on structural MRI. Within those, predictions based on grey-matter density and cerebrospinal fluid showed the highest agreement with a Kappa value of 0.4 (considered a 'moderate' agreement⁷⁸). All other modality comparisons exhibited relatively little agreement with Kappa values ≤ 0.25 . Across all modalities, predictions based on nucleus accumbens connectivity showed the least agreement with any other modality.

Multimodal diagnostic classification performs better than each individual modality. To combine modalities for multimodal classification, we trained individual classifiers on each modality separately, computed predictions based on each modality, and then combined these predictions in the form of a weighted ensemble vote (Fig. 1). A number of details of this analytic framework deserve highlighting. First, classifiers were optimized for each modality individually, both with respect to the classifier type (SVM or WeiRD) and in terms of a modality-specific regularization parameter (C in the case of SVM). Second, instead of relying on discrete predictions, continuous decision values were computed for each modality, which enabled a more fine-grained ensemble decision. Third, on top of these intrinsic classifier-based decision values, an external weighting matrix across modalities was learned from the training data, further refining the ensemble vote.

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

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

To assess the significance of each individual modality for multimodal classification, we computed classification accuracies while tentatively excluding each modality once. The results (Table 4) showed that the accuracy dropped in each case, indicating that each modality was important for overall multimodal performance. Not surprisingly, the sharpest drop in performance was observed when excluding the best individual modality, grey-matter density (−9.6%). In particular, grey-matter density was largely responsible for the high sensitivity of

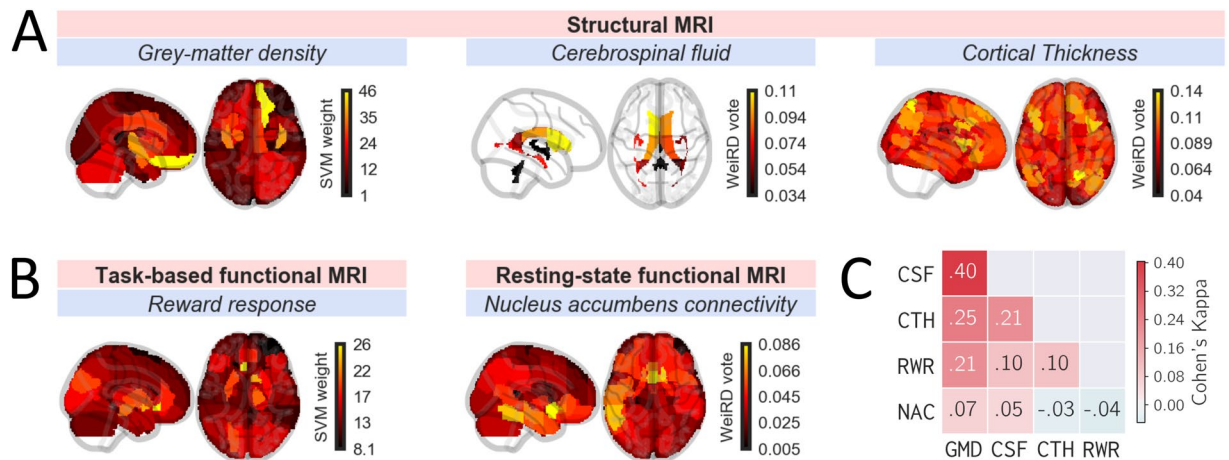


Figure 2. Unimodal classification. Feature importances of (A) structural and (B) functional neuroimaging modalities. Depicted are 2-*d* projections (‘glass brains’) of feature importances along the *x*- and *z*-axis. Feature importances represent SVM weights (grey-matter density, reward response) or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

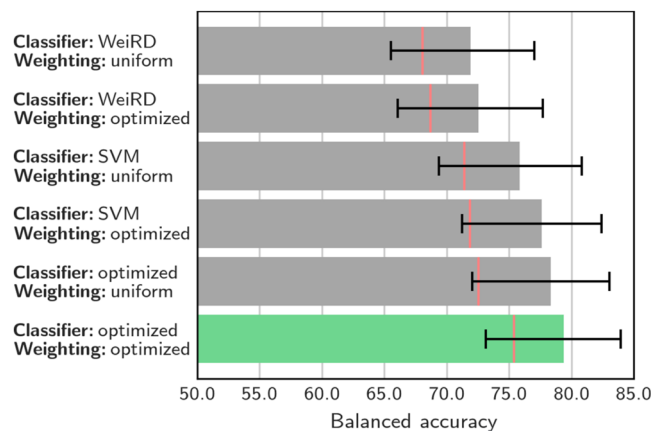


Figure 3. Multimodal classification. Balanced accuracy for classification schemes based on different classifier configurations (SVM, WeiRD or optimized between SVM and WeiRD) and with uniform (i.e., all weights set to 1) or optimized weighting of modalities. Optimizing both classifiers and weighting yielded the best performance (highlighted in green). Red lines indicate the balanced accuracy when ensemble prediction was based on discrete “control” and “patient” judgements instead of continuous decision values. Error bars represent the 95% posterior probability interval⁶⁷.

Acc. (Δ)	Sens. (Δ)	Spec. (Δ)
69.7 (−9.6)	59.2 (−18.1)	80.2 (−1.0)
75.3 (−4.0)	76.7 (−0.6)	74.0 (−7.3)
78.1 (−1.2)	75.8 (−1.5)	80.4 (−0.8)
77.0 (−2.3)	75.6 (−1.7)	78.4 (−2.9)
77.3 (−2.0)	77.3 (0.0)	77.3 (−3.9)

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

multimodal classification, evidenced by a substantial drop of sensitivity when excluded (−18.1%). By contrast, the sharpest drop in specificity was found when excluding cerebrospinal fluid (−7.3%). Taken together, these results show that the success of multimodal classification was based on the combined information of all modalities such that no single modality was effectively redundant.

Discussion

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

Our initial unimodal analysis of feature importances showed that informative features were primarily located in (orbito-)prefrontal and cingulate brain regions as well as in the nucleus accumbens (for basic reward responses). These foci are broadly in line with previous investigations into brain structure and function in alcohol dependence, which identified executive and reward networks as major neural circuits that are affected in patients⁷⁵. Thus, our unimodal structural and functional data show diagnostic characteristics well in line with previous results. Despite the overlap of affected brain structures, none of these considered modalities was redundant: multimodal classification accuracy was reduced for each modality that was tentatively excluded. Together, these exploratory analyses into individual modalities thus attest to 1) plausible unimodal between-group effects and 2) a sensible selection of modalities that leveraged on non-redundant sources of information.

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

Although the improvement in overall accuracy is modest at 2.7%, one important qualification is that the comparison to the best of several modalities is necessarily unfair due to selection bias; that is, it is likely that the accuracy of the best modality is – to some degree – inflated by noise. Besides this ‘double dipping’⁷⁹ bias, even at 2.7% our observed benefit of multimodal classification is noteworthy, as many previous studies have failed to find any advantage of combining multiple modalities^{9,10,80,81}. There is currently no consensus on *why* an effective combination of multiple MRI modalities appears to be a surprisingly intricate task. But clearly, the success of multimodal classification must depend on how individual modalities complement each other: if individual modalities perfectly agree with each other, they are effectively redundant and thus no multimodal benefit is expected; vice versa, if predictions of individual modalities are completely uncorrelated, there is no positive cumulative effect through the combination of modalities. In the present case, the agreement of predictions between different modalities was quite variable overall but tended towards low inter-modality reliabilities (Cohen’s Kappa scores up to 0.4), which reflects the moderate benefit of multimodal classification.

At a methodological level, we found a number of analytic choices to be critical for the integration of multiple modalities. First, as data from different modalities will naturally have different structural properties, it is likely that there is no single classifier type that fits all modalities. Yet, to our knowledge, while multimodal investigations often *compare* classifier types^{9,11,82}, they do so by applying the same classifier to all modalities, i.e. they do not optimize the classifier type in a modality-specific manner. Here we found that optimizing the classifier type for each modality was superior to using either SVM or WeIRD uniformly across modalities. Second, applying a weight matrix learned from training data to the ensemble prediction likewise increased the performance compared to unweighted integration. Although this particular aspect has been considered previously (e.g. by means of logistic regression on predictions of individual modalities¹¹), our results corroborate the importance of this analytic step.

Third, the biggest gain in accuracy was owed to using continuous classifier decision values instead of discrete (binary) predictions. The benefit of decision values can be explained by the fact that for both SVM and WeIRD, decision values reflected the certainty of the classifier. Thus, more certain predictions factor in more strongly into the ensemble prediction and thereby improve the overall accuracy. Indeed, we have previously shown that valuable information is contained in unthresholded decision values of classifiers applied to neuroimaging data⁶³. On the basis of our results we thus recommend to avoid discretizing modality-specific predictions and instead to utilize fine-grained information contained in classifier decision values.

Despite these methodological insights about combining multiple neuroimaging modalities from both structural and functional MRI, considering the fact that fMRI scanning was more time-consuming and involved more elaborate data analysis, the achieved improvement through fMRI must be debated in view of clinical practicability and a cost-benefit analysis. In our view, several factors are important for fMRI to become part of a clinically realistic diagnostic imaging battery. First, an accuracy net gain in the order of a few percent is relatively small, given the additional effort of instructing and conducting a functional scan in addition to a structural scan. Thus, to justify task-based and resting-state fMRI measures, these must either be more sensitive – or 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. fMRIprep⁸³), 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^{84–86}. 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 imaging¹⁶ 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 share major aspects of the methodology, including the use of the weighted robust distance classifier⁶² and 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^{75,87}. 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 behaviours⁷¹. 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 and may as well be informative for diagnostic classification⁸⁸. These large-scale networks were not considered in the present study and thus deserve further research in the context of diagnostic neuroimaging-based classification. Fourth, based on the cross-sectional design in this study one cannot infer whether the neurobiological differences utilized by our classifier are causes, state markers or consequences of alcohol dependence (although see preliminary evidence for the latter possibility in Supplementary Fig. 2). Thus, our classifier provides no mechanistic insight into the pathogenesis of alcohol dependence. For instance, based on previous findings it is possible that group differences in brain structure may be predominantly consequences of severe alcohol abuse, as these changes partially reverse during abstinence^{89–91}; differences in brain function may, in turn, predispose for addiction or relapse from addiction^{30,92}. From a strict machine-learning point of view, one may be tempted to treat the issue of mechanistic insight as secondary. However, although neural features that characterize a predisposition are particularly valuable due to their prognostic potential, for diagnosis they bear the risk of misclassifying healthy individuals that did not develop AD despite having a disadvantageous predisposition. Patients which developed AD without such a predisposition could likewise be misclassified due to these features. For this reason, the ‘chicken or egg’ causality dilemma is indeed relevant for machine-learning-based diagnosis and should be investigated in future longitudinal studies.

Finally, in view of modern frameworks of psychiatric disease such as the Research Domain Criteria (RDoC) project, in which psychiatric phenotypes are defined as “spanning the range from normal to abnormal”^{93,94}, the present approach could be readily adapted to predict dimensional markers of disease. Analogous to the idea of a weighted voting scheme across classifiers applied here for the case of binary prediction, multiple modality-specific regression models would be trained and their outputs combined to form a continuous ensemble prediction. The ensemble prediction would likewise be based on a weighted (and normalized) sum of individual predictions. Conceivable dimensional markers for the case of alcohol dependence are the magnitude of craving in acute addiction⁹⁵, biological markers such as serum levels of carbohydrate deficient transferrin (CDT) and gamma glutamyltransferase (GGT)⁹⁶, or scores of clinical questionnaires such as the Alcohol Use Disorders Identification Test (AUDIT⁹⁷). Overall, we conclude that the combination of multiple neuroimaging modalities is able to moderately improve the accuracy of machine-learning-based diagnostic classification in alcohol dependence. Our results allow us to make several methodological recommendations for the exploitation and integration of different modalities with the goal to compute optimal ensemble predictions, thereby paving the way towards more effective multimodal neuroimaging classifiers. Yet, at present, given the strong predictive performance of grey-matter density alone and taking a cost-benefit analyses into account, we currently recommend to focus on structural MRI for the diagnostic classification of alcohol dependence.

Data availability

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

Received: 26 July 2019; Accepted: 19 December 2019;

Published online: 15 January 2020

References

- Liu, S. *et al.* Multimodal neuroimaging computing: a review of the applications in neuropsychiatric disorders. *Brain Informatics* **2**, 167–180 (2015).
- Perrin, R. J., Fagan, A. M. & Holtzman, D. M. Multimodal techniques for diagnosis and prognosis of Alzheimer’s disease. *Nature* **461**, 916–922 (2009).
- Calhoun, V. D. & Sui, J. Multimodal Fusion of Brain Imaging. *Data: A Key to Finding the Missing Link(s) in Complex Mental Illness. Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **1**, 230–244 (2016).
- Santelmann, H., Franklin, J., Bußhoff, J. & Baethge, C. Test-retest reliability of schizoaffective disorder compared with schizophrenia, bipolar disorder, and unipolar depression—a systematic review and meta-analysis. *Bipolar Disord.* **17**, 753–768 (2015).
- Freedman, R. *et al.* The initial field trials of DSM-5: New blooms and old thorns. *Am. J. Psychiatry* **170**, 1–5 (2013).
- Wardenaar, K. J. & de Jonge, P. Diagnostic heterogeneity in psychiatry: Towards an empirical solution. *BMC Med.* **11**, 2–4 (2013).
- Jablensky, A. Psychiatric classifications: Validity and utility. *World Psychiatry* **15**, 26–31 (2016).
- Callard, F. Psychiatric diagnosis: The indispensability of ambivalence. *J. Med. Ethics* **40**, 526–530 (2014).
- Ritter, K. *et al.* Multimodal prediction of conversion to Alzheimer based on incomplete biomarkers*This work was supported by the Bernstein Computational Program of the German Federal Ministry of Education and Research (01GQ1001C, 01GQ0851, GRK 1589/1), the European Regiona. *Alzheimer’s Dement. Diagnosis, Assess. Dis. Monit.* **1**, 206–215 (2015).
- Shaffer, J. L. *et al.* Predicting Cognitive Decline in Subjects at Risk for Alzheimer Disease by Using Combined. *Radiology* **266**, 583–591 (2013).

11. Young, J. *et al.* Accurate multimodal probabilistic prediction of conversion to Alzheimer's disease in patients with mild cognitive impairment. *NeuroImage Clin.* **2**, 735–745 (2013).
12. Klöppel, S. *et al.* Automatic classification of MR scans in Alzheimer's disease. *Brain* **131**, 681–689 (2008).
13. Hackmack, K. *et al.* Can we overcome the 'clinico-radiological paradox' in multiple sclerosis? *J. Neurol.* **259**, 2151–2160 (2012).
14. Weygandt, M. *et al.* MRI pattern recognition in multiple sclerosis normal-appearing brain areas. *PLoS ONE* **6**, (2011).
15. Harper, C. & Matsumoto, I. Ethanol and brain damage. *Curr. Opin. Pharmacol.* **5**, 73–78 (2005).
16. Guggenmos, M. *et al.* Decoding diagnosis and lifetime consumption in alcohol dependence from grey-matter pattern information. *Acta Psychiatr. Scand.* **137**, 252–262 (2018).
17. Jernigan, T. L. *et al.* Reduced cerebral grey matter observed in alcoholics using magnetic resonance imaging. *Alcohol. Clin. Exp. Res.* **15**, 418–427 (1991).
18. Pfefferbaum, A. *et al.* Brain gray and white matter volume loss accelerates with aging in chronic alcoholics: a quantitative MRI study. *Alcohol. Clin. Experimental Res.* **16**, 1078–89 (1992).
19. Fox, J. H., Ramsey, R. G., Huckman, M. S. & Proske, A. E. Cerebral Ventricular Enlargement Chronic. *JAMA* **4**, 365–368 (1976).
20. Ron, M. A. *The alcoholic brain: CT scan and psychological findings. Psychological Medicine. Monograph Supplement* **3**, (1983).
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).
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).
23. Wrase, J. *et al.* Dysfunction of reward processing correlates with alcohol craving in detoxified alcoholics. *Neuroimage* **35**, 787–794 (2007).
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).
25. Tapert, S. F. *et al.* Neural response to alcohol stimuli in adolescents with alcohol use disorder. **60**, 727–735 (2003).
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).
27. Ruskin, P. *et al.* Reliability and acceptability of psychiatric diagnosis via telecommunication and audiovisual technology. *Psychiatr. Serv.* **49**, 1086–1088 (1998).
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).
29. Guggenmos, M. *et al.* Quantitative neurobiological evidence for accelerated brain aging in alcohol dependence. *Transl. Psychiatry* **7**, 1279 (2017).
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).
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).
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).
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).
34. Garbusow, M. *et al.* Pavlovian-to-Instrumental Transfer in Alcohol Dependence: A Pilot Study. *Neuropsychobiology* **70**, 111–121 (2014).
35. Garbusow, M., Sebold, M., Beck, A. & Heinz, A. Too Difficult to Stop: Mechanisms Facilitating Relapse in Alcohol Dependence. *Neuropsychobiology* **70**, 103–110 (2014).
36. Wittchen, H.-U. & Pfister, H. *DIA-X-Interviews: Manual für Screening-Verfahren und Interview; Interviewheft.* (Swets & Zeitlinger, 1997).
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).
38. Oldfield, R. C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* **9**, 97–113 (1971).
39. Kotz, D., Böckmann, M. & Kastaun, S. Nutzung von Tabak und E-Zigaretten sowie Methoden zur Tabakentwöhnung in Deutschland. *Dtsch. Arztebl. Int.* **115**, 235–242 (2018).
40. Fortier, C. B. *et al.* Reduced Cortical Thickness in Abstinent Alcoholics and Association with Alcoholic Behavior. *Alcohol. Clin. Exp. Res.* **35**, 2193–2201 (2011).
41. Mashhoon, Y. *et al.* Binge alcohol consumption in emerging adults: Anterior cingulate cortical 'thinness' is associated with alcohol use patterns. *Alcohol. Clin. Exp. Res.* **38**, 1955–1964 (2014).
42. Momenan, R. *et al.* Effects of alcohol dependence on cortical thickness as determined by magnetic resonance imaging. *Psychiatry Res. - Neuroimaging* **204**, 101–111 (2012).
43. Faria, A. V. *et al.* Atlas-based analysis of resting-state functional connectivity: Evaluation for reproducibility and multi-modal anatomy–function correlation studies. *Neuroimage* **61**, 613–621 (2012).
44. Dale, A. M., Fischl, B. & Sereno, M. I. Cortical Surface-Based Analysis. I. Segmentation and Surface Reconstruction. *Neuroimage* **9**, 179–194 (1999).
45. Fischl, B. & Dale, A. M. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl. Acad. Sci.* **97**, 11050–11055 (2000).
46. Fischl, B., Sereno, M. I. & Dale, A. M. Cortical Surface-Based Analysis. II: Inflation, Flattening, and a Surface-Based Coordinate System. *Neuroimage* **9**, 195–207 (1999).
47. Fischl, B., Sereno, M. I., Tootell, R. B. H. & Dale, A. M. High-resolution inter-subject averaging and a surface-based coordinate system. *Hum. Brain Mapp.* **8**, 272–284 (1999).
48. Fischl, B., Liu, A. & Dale, A. M. Automated manifold surgery: Constructing geometrically accurate and topologically correct models of the human cerebral cortex. *IEEE Trans. Med. Imaging* **20**, 70–80 (2001).
49. Fischl, B. *et al.* Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341–355 (2002).
50. Fischl, B. *et al.* Automatically Parcellating the Human Cerebral Cortex. *Cereb. Cortex* **14**, 11–22 (2004).
51. Fischl, B. *et al.* Sequence-independent segmentation of magnetic resonance images. *Neuroimage* **23**, 69–84 (2004).
52. Han, X. *et al.* Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer. *Neuroimage* **32**, 180–194 (2006).
53. Jovicich, J. *et al.* Reliability in multi-site structural MRI studies: Effects of gradient non-linearity correction on phantom and human data. *Neuroimage* **30**, 436–443 (2006).
54. Ségonne, F. *et al.* A hybrid approach to the skull stripping problem in MRI. *Neuroimage* **22**, 1060–1075 (2004).
55. Reuter, M., Schmansky, N. J., Rosas, H. D. & Fischl, B. Within-subject template estimation for unbiased longitudinal image analysis. *Neuroimage* **61**, 1402–1418 (2012).
56. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
57. Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P. & Dolan, R. J. Model-Based Influences on Humans' Choices and Striatal Prediction Errors. *Neuron* **69**, 1204–1215 (2011).

58. Pruim, R. H. R. *et al.* ICA-AROMA: A robust ICA-based strategy for removing motion artifacts from fMRI data. *Neuroimage* **112**, 267–277 (2015).
59. Pruim, R. H. R., Mennes, M., Buitelaar, J. K. & Beckmann, C. F. Evaluation of ICA-AROMA and alternative strategies for motion artifact removal in resting state fMRI. *Neuroimage* **112**, 278–287 (2015).
60. Hsu, C.-W., Chang, C.-C. & Lin, C.-J. *A Practical Guide to Support Vector Classification*. <https://doi.org/10.1177/02632760022050997> (2016).
61. Cortes, C. & Vapnik, V. Support-Vector Networks. *Mach. Learn.* **20**, 273–297 (1995).
62. Guggenmos, M., Schmack, K. & Sterzer, P. WeiRD - a fast and performant multivoxel pattern classifier. *2016 Int. Work. Pattern Recognit. Neuroimaging 1–4* <https://doi.org/10.1109/PRNI.2016.7552349> (2016).
63. Guggenmos, M., Sterzer, P. & Cichy, R. M. Multivariate pattern analysis for MEG: a comparison of dissimilarity measures. *Neuroimage* **173**, 434–447 (2018).
64. Zhou, L., Wang, L., Liu, L., Ogunbona, P. O. & Shen, D. Support Vector Machines for neuroimage analysis: Interpretation from discrimination. In *Support Vector Machines Applications* 191–220 (Springer, 2014).
65. Chang, C. & Lin, C. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* **2**, 27:1–27:27 (2011).
66. Luntz, A. & Brailovsky, V. On estimation of characters obtained in statistical procedure of recognition. *Tech. Kibern.* **3**, 6–12 (1969).
67. Brodersen, K. H., Ong, C. S., Stephan, K. E. & Buhmann, J. M. The balanced accuracy and its posterior distribution. *Proc. - Int. Conf. Pattern Recognit.* 3121–3124 <https://doi.org/10.1109/ICPR.2010.764> (2010).
68. Dwyer, D. B., Falkai, P. & Koutsouleris, N. Machine Learning Approaches for Clinical Psychology and Psychiatry. *Ssrn* 1–28 <https://doi.org/10.1146/annurev-clinpsy-032816-045037> (2018).
69. Haxby, J. V., Connolly, A. C. & Guntupalli, J. S. Decoding Neural Representational Spaces Using Multivariate Pattern Analysis. *Annu. Rev. Neurosci.* 435–456 <https://doi.org/10.1146/annurev-neuro-062012-170325> (2014).
70. Jamalabadi, H., Alizadeh, S., Schönauer, M., Leibold, C. & Gais, S. Classification based hypothesis testing in neuroscience: Below-chance level classification rates and overlooked statistical properties of linear parametric classifiers. *Hum. Brain Mapp.* **37**, 1842–1855 (2016).
71. Goldstein, R. Z. & Volkow, N. D. Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nat. Rev. Neurosci.* **12**, 652–669 (2012).
72. Abernathy, K., Chandler, L. J. & Woodward, J. J. Alcohol and the Prefrontal Cortex. *Int. Rev. Neurobiol.* **91**, 289–320 (2010).
73. Perry, J. L. *et al.* Prefrontal cortex and drug abuse vulnerability: Translation to prevention and treatment interventions. *Brain Res. Rev.* **65**, 124–149 (2011).
74. Ikemoto, S. & Panksepp, J. The role of nucleus accumbens dopamine in motivated behavior: A unifying interpretation with special reference to reward-seeking. *Brain Res. Rev.* **31**, 6–41 (1999).
75. Volkow, N. D., Wang, G.-J., Fowler, J. S. & Tomasi, D. Addiction Circuitry in the Human Brain. *Annu. Rev. Pharmacol. Toxicol.* **52**, 321–336 (2012).
76. Goldstein, R. Z. & Volkow, N. D. Drug addiction and its underlying neurobiological basis: Neuroimaging evidence for the involvement of the frontal cortex. *Am. J. Psychiatry* **159**, 1642–1652 (2002).
77. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **20**, 37–46 (1960).
78. Landis, R. J. & Koch, G. G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **33**, 159–174 (1977).
79. Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F. & Baker, C. I. Circular analysis in systems neuroscience: the dangers of double dipping. *Nat. Neurosci.* **12**, 535–40 (2009).
80. Cui, Y. *et al.* Identification of conversion from mild cognitive impairment to alzheimer’s disease using multivariate predictors. *PLoS One* **6**, 2–11 (2011).
81. Pettersson-Yeo, W. *et al.* An empirical comparison of different approaches for combining multimodal neuroimaging data with support vector machine. *Front. Neurosci.* **8**, (2014).
82. Ray, B. *et al.* Information content and analysis methods for multi-modal high-throughput biomedical data. *Sci. Rep.* **4**, 1–10 (2014).
83. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *bioRxiv* <https://doi.org/10.1101/306951> (2018).
84. Agartz, I., Shoaf, S., Rawlings, R. R., Momenan, R. & Hommer, D. W. CSF monoamine metabolites and MRI brain volumes in alcohol dependence. *Psychiatry Res. - Neuroimaging* **122**, 21–35 (2003).
85. Hommer, D. W., Momenan, R., Kaiser, E. & Rawlings, R. R. Evidence for a gender-related effect of alcoholism on brain volumes. *Am. J. Psychiatry* **158**, 198–204 (2001).
86. Schweinsburg, B. C. *et al.* Effects of alcoholism and gender on brain metabolism. *Am. J. Psychiatry* **160**, 1180–1183 (2003).
87. Sullivan, E. V. & Pfefferbaum, A. Neuropsychology and neuroimaging studies in alcohol-dependence 30 ans de recherche en neuropsychologie et en imagerie cérébrale sur l’alcool-dépendance. **5**, 187–199 (2013).
88. Fede, S. J., Grodin, E. N., Dean, S. F., Diazgranados, N. & Momenan, R. Resting state connectivity best predicts alcohol use severity in moderate to heavy alcohol users. *NeuroImage Clin.* **22**, 101782 (2019).
89. Demirakca, T. *et al.* Effects of Alcoholism and Continued Abstinence on Brain Volumes in Both Genders. *Alcohol. Clin. Exp. Res.* **35**, 1678–1685 (2011).
90. Cardenas, V. A., Studholme, C., Gazdzinski, S., Durazzo, T. C. & Meyerhoff, D. J. Deformation-based morphometry of brain changes in alcohol dependence and abstinence. *Neuroimage* **34**, 879–887 (2007).
91. Sullivan, E. V., Rosenbloom, M. J., Lim, K. O. & Pfefferbaum, A. Longitudinal changes in cognition, gait, and balance in abstinent and relapsed alcoholic men: Relationships to changes in brain structure. *Neuropsychology* **14**, 178–188 (2000).
92. Whelan, R. *et al.* Neuropsychosocial profiles of current and future adolescent alcohol misusers. *Nature* **512**, 185–189 (2014).
93. National Institute of Mental Health. Research Domain Criteria (RDoC). (2013).
94. Yücel, M. *et al.* A transdiagnostic dimensional approach towards a neuropsychological assessment for addiction: an international Delphi consensus study. *Addiction* **114**, 1095–1109 (2019).
95. Sayette, M. A. *et al.* The measurement of drug craving. *Addiction* **95**, 189–210 (2000).
96. Conigrave, K. M. *et al.* CDT, GGT, and AST as markers of alcohol use: The WHO/ISBRA Collaborative Project. *Alcohol. Clin. Exp. Res.* **26**, 332–339 (2002).
97. Saunders, J. B. *et al.* Development of the Alcohol Use Disorders Identification Test (AUDIT). *Addiction* 791–804 (1993).
98. Schmidt, L., Gastpar, M., Falkai, P. & Gabel, W. Evidenzbasierte Suchtmedizin. In *Substanzbezogene Störungen* (Deutscher Ärzteverlag, 2006).
99. Skinner, H. & Horn, J. *Alcohol Dependence Scale (ADS): User’s Guide*. (1984).
100. Mann, K. & Ackermann, K. Die OCDS-G: Psychometrische Kennwerte der deutschen Version der Obsessive Compulsive Drinking Scale [The OCDS-G: Psychometric Characteristics of the German version of the Obsessive Compulsive Drinking Scale]. *Sucht* **46**, 90–100 (2000).
101. Meule, A., Vögele, C. & Kübler, A. Psychometrische evaluation der Deutschen Barratt Impulsiveness scale - Kurzversion (BIS-15). *Diagnostica* **57**, 126–133 (2011).
102. *Army Individual Test Battery. Manual of directions and scoring.* War Department, Adjutant General’s Office. (1944).
103. Wechsler, D. *WAIS-III, Wechsler Adult Intelligence Scale: Administration and Scoring Manual*. (1997).
104. Lehl, S. *Mehrfachwahl-Wortschatz-Intelligenztest MWT-B*. (Spitta, 2005).
105. Morris, J. C. *et al.* The Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) - Part I. Clinical and neuropsychological assessment of Alzheimer’s disease. *Neurology* **39**, 1159–65 (1989).

Acknowledgements

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, WA 1539/7–1, 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 contributions

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.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-019-56923-9>.

Correspondence and requests for materials should be addressed to M.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020