nature mental health

Analysis

https://doi.org/10.1038/s44220-023-00173-2

Brain-based classification of youth with anxiety disorders: transdiagnostic examinations within the ENIGMA-Anxiety database using machine learning

Received: 14 March 2023

Accepted: 26 October 2023

Published online: 10 January 2024

Check for updates

A list of authors and their affiliations appears at the end of the paper

Neuroanatomical findings on youth anxiety disorders are notoriously difficult to replicate, small in effect size and have limited clinical relevance. These concerns have prompted a paradigm shift toward highly powered (that is, big data) individual-level inferences, which are data driven, transdiagnostic and neurobiologically informed. Here we built and validated supervised neuroanatomical machine learning models for individual-level inferences, using a case–control design and the largest known neuroimaging database on youth anxiety disorders: the ENIGMA-Anxiety Consortium (N = 3,343; age = 10–25 years; global sites = 32). Modest, yet robust, brain-based classifications were achieved for specific anxiety disorders (panic disorder), but also transdiagnostically for all anxiety disorders when patients were subgrouped according to their sex, medication status and symptom severity (area under the receiver operating characteristic curve, 0.59–0.63). Classifications were driven by neuroanatomical features (cortical thickness, cortical surface area and subcortical volumes) in fronto-striato-limbic and temporoparietal regions. This benchmark study within a large, heterogeneous and multisite sample of youth with anxiety disorders reveals that only modest classification performances can be realistically achieved with machine learning using neuroanatomical data.

Anxiety disorders are the most prevalent mental disorders among youth, with a lifetime prevalence estimate of up to 30% (refs. 1,2). Most anxiety disorders develop during the critical transition from adolescence to young adulthood (10–25 years), affecting millions of youth worldwide and causing enormous emotional, societal and economic burden[3,4]. Critically, the COVID-19 pandemic has further exacerbated this alarming trend, with some experts now even talking about a 'lost generation' of youth[5]. Psychopathology is less differentiated in youth than in adults, and thus even less compatible with traditional diagnostic nosology, which further complicates the situation[6–8]. Despite these concerns, the underlying neurobiology of anxiety disorders in youth remains elusive, making it difficult to pinpoint robust biomarkers and formulate or test neurobiologically informed treatment and/or prevention strategies[4,9].

Neuroimaging studies point to anomalies in fronto-striato-limbic brain circuits and additional sensory regions in youth with anxiety disorders[6,9,10] that collectively affect the perception, processing and modulation of emotionally salient information (Supplementary Fig. 1). While promising, these neuroimaging findings are often difficult to replicate and small in effect size, with marginal clinical relevance[4,9,11–14]. These issues may reflect limitations of our currently dominant analytic approach, which favors disorder-specific analyses rather than transdiagnostic ones, often within underpowered samples and with the use of traditional mass univariate analyses that preclude massively multivariate and individual-level analytics[11–15]. These limitations have prompted a shift in the field toward highly powered (that is, big data) individual-level inferences that are data driven, transdiagnostic

e-mail: m.aghajani@fsw.leidenuniv.nl

https://doi.org/10.1038/s44220-023-00173-2

Table 1 | Demographic and clinical characteristics for included patients and controls of the three ENIGMA-Anxiety Working Groups

	ENIGMA-Anxiety Working Group									Transdiagnostic(32 sites)		
	PD(12 sites)			GAD(16 sites)			SAD(13 sites)					
Characteristic	**Patients (*n*=112)**	**Controls (*n*=813)**	***P***	**Patients (*n*=465)**	**Controls (*n*=1,084)**	***P***	**Patients (*n*=259)**	**Controls (*n*=610)**	***P***	**Patients (*n*=823)[e]**	**Controls (*n*=1,969)[e]**	***P***
Age, years	21.68±2.78	21.73±3.01	0.86	18.39±4.28	16.06±4.35	<0.001	21.36±2.45	22.17±2.30	<0.001	19.79±3.90	18.79±4.75	<0.001
Male/female sex, *n*	25/87	275/538	0.02	148/317	495/589	<0.001	78/181	188/422	0.90	248/575	773/1196	<0.001
STAI-T score[a]	50.07±12.25	–		45.25±11.14	–		52.12±11.08	–		48.10±11.72	–	
Medication, *n* (%)[b]	37 (35.58)	–		91 (20.13)	–		35 (13.57)	–		160 (19.98)	–	
Antidepressants	31 (30.39)	–		59 (13.05)	–		30 (12.35)	–		118 (14.96)	–	
Benzodiazepines	6 (8.22)	–		30 (6.80)	–		0 (0)	–		36 (4.59)	–	
Antipsychotics	9 (10.98)	–		25 (5.67)	–		4 (1.80)	–		37 (5.02)	–	
Comorbid anxiety, *n* (%)[c]												
PD	–	–		38 (9.43)	–		8 (3.09)	–		36 (5.55)	–	
GAD	38 (34.86)	–		–	–		12 (4.63)	–		46 (12.78)	–	
SAD	32 (28.57)	–		223 (49.34)	–		–	–		223 (43.98)	–	
Mixed[d]	20 (18.35)	–		23 (5.71)	–		1 (0.39)	–		41 (5.41)	–	

For age and STAI-T score, values are mean ± s.d. [a]Available data on STAI-T: PD=88, GAD=306 and SAD=182. [b]Available data on psychotropic medication use at the time of scan: PD=104, GAD=452 and SAD=258. [c]Available data on comorbid anxiety disorders: PD=109, GAD=403 and SAD=259. Assigned primary anxiety disorder diagnoses correspond to respective working groups. [d]Patients diagnosed with all three anxiety disorders. [e]Participants present in more than one working group were excluded from transdiagnostic classifications to avoid the use of duplicated data entries.

and informed by objective neurobiological measures[11,15]. This shift is anticipated to provide mechanistic neurobiological insights that would help pinpoint reliable biomarkers and potential therapeutic targets in individual subjects[11,15].

The application of machine learning (ML) may be particularly useful for this endeavor[11,15,16]. ML is well suited to dealing with high-dimensional data in a data-driven manner, extracting complex multivariate patterns that best predict individual-level clinical outcomes[11,15,16]. Previous brain-based ML classification studies of anxiety disorders show that neuroimaging data can distinguish patients from controls with varying success and accuracies up to 90% (refs. 17–24). Although these studies are encouraging, most are limited by their use of small, single-site samples, so it is unclear whether results generalize to data from other sites with different demographic (for example, age and/or sex distribution), clinical (for example, medication, symptom severity and comorbidity) or technical (for example, scanner, acquisition protocols and diagnostic assessment) characteristics. To overcome these limitations, large multisite collaborations have begun to pool neuroimaging and clinical datasets for coordinated analyses, wherein all data are preprocessed and analyzed according to harmonized and standardized protocols. The Enhancing Neuro-Imaging Genetics through Meta-Analysis (ENIGMA) Consortium is such an initiative with a massive global reach[25,26], and was hence used for this study. The large-scale, multisite ENIGMA datasets offer ecologically valid and clinically representative information, which ML algorithms can exploit to identify multivariate patterns generalizable to the majority of patients, in a realistic fashion[27–29]. Although increasing sample size typically benefits ML performance and generalizability, the use of retrospectively pooled multisite data may also complicate performance due to the substantial heterogeneity (for example, sample characteristics and methodology) that is inherently introduced[27–30].On the other hand, a multisite design may provide more realistic estimates of performance and can be used to explicitly test the robustness of predictive models, which is a necessary prerequisite for implementation into routine care.

In this study, we built and validated neuroanatomical magnetic resonance imaging (MRI)-based ML models, using the largest multisite sample of neuroimaging data with young anxiety patients and healthy controls (HC) from across the globe (N = 3,343 from 32 sites). The sample was aggregated from three subgroups of the ENIGMA-Anxiety Working Group comprising the most prevalent youth anxiety disorders: panic disorder (PD), generalized anxiety disorder (GAD) and social anxiety disorder (SAD) (Table 1)[10]. We focused on individuals between 10 and 25 years old, thus capturing the critical transition from adolescence to young adulthood in which these anxiety disorders first develop[3,4,31]. ML classifiers were trained on MRI-derived cortical and subcortical gray matter features (regional cortical thickness (CT)/surface area (CSA) and subcortical volume (SCV)) to classify patients versus HC for each anxiety disorder separately (disorder specific) and across anxiety disorders (hereafter referred to as transdiagnostic classification). Our hypothesis was that the classifiers would correctly distinguish patients versus HC above chance level. We aimed for entirely brain-based classifications, with additional exploratory analyses examining the influence of sex, medication use, symptom severity and depression comorbidity on brain-based classification performance.

Results

Primary analyses

Case–control classification performances are summarized in Fig. 1a, and a complete overview of results (area under the receiver operating characteristic curve (AUC), balanced accuracy, sensitivity, specificity and permutation *P*values) is provided in Supplementary Table 2. The best classification performance was obtained for PD versus HC (112 patients with PD and 813 HC; 12 sites) with an average AUC of 0.62 (P = 0.027). None of the other disorder-specific classification performances were significantly different from chance: GAD classification (465 participants with GAD and 1,084 HC; 16 sites; average AUC = 0.55, P = 0.605), SAD classification (259 participants with SAD and 610 HC; 13 sites; average AUC = 0.57, P = 0.32) and transdiagnostic classification (823 participants with any anxiety disorder and 1,969 HC; 32 sites; average AUC = 0.56, P = 0.093). Classification performances obtained using leave-one-site-out cross-validation (LOSO-CV) were similar to those obtained with repeated fivefold CV, in which only PD classification performed significantly above chance level (average AUC = 0.63, P = 0.003), whereas the SAD, GAD and transdiagnostic classifications failed to surpass chance level (Supplementary Table 3).

https://doi.org/10.1038/s44220-023-00173-2

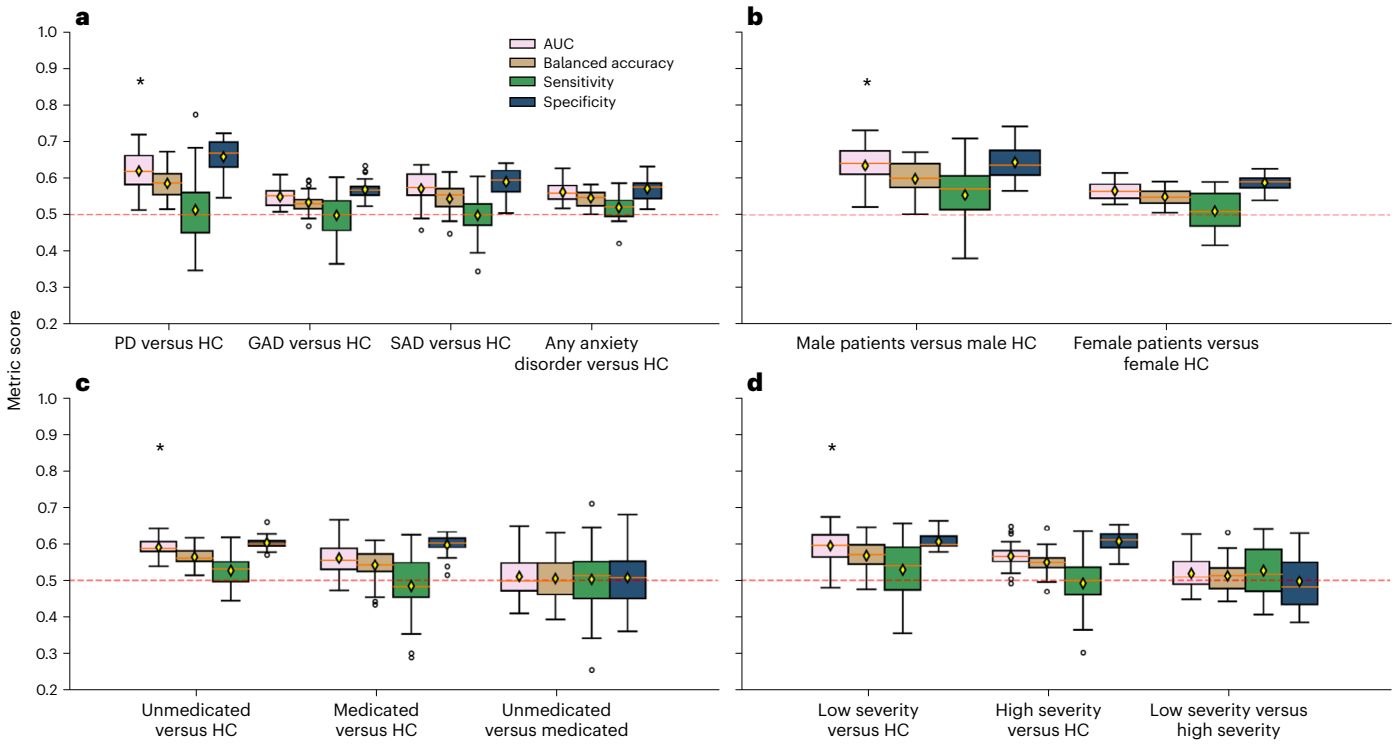

Fig. 1 | Classification performances for each working group and transdiagnostic classification performance across working groups. The box plots summarize scores obtained across the repeated stratified *K*-fold CV folds; the yellow diamonds indicate mean scores, circles indicate outliers, asterisks depict significance as assessed by label permutation tests and the dashed line represents chance-level performance. The bounds of the upper whiskers represent the maximum, the bounds of the lower whiskers represent the minimum, the center box represents the interquartile range and the mean is represented by the middle horizontal lines. Exact *P* values, along with all classification metrics, are provided in Supplementary Table 2. Sample size per classification task: for PD versus HC, $n = 112$ versus 813; for GAD versus HC, $n = 465$ versus 1,084; for SAD versus HC, $n = 259$ versus 610; for any anxiety disorder versus HC, $n = 823$ versus 1,969; for male patients versus male HC, $n = 167$ versus 678; for female patients versus female HC, $n = 524$ versus 1,133; for unmedicated patients versus HC, $n = 641$ versus 1,969; for medicated patients versus HC, $n = 160$ versus 1,307; for unmedicated versus medicated patients, $n = 251$ versus 160; for low-severity anxiety versus HC, $n = 299$ versus 1,422; for high-severity anxiety versus HC, $n = 272$ versus 1,449; and for low- versus high-severity anxiety, $n = 299$ versus 270. **a**, Classification performances for each working group. **b**–**d**, Transdiagnostic classification performances across working groups.

Exploratory subgroup analyses

Sex. We assessed transdiagnostic case–control classification (across ENIGMA-Anxiety subgroups) separately for male and female participants. The performances obtained are summarized in Fig. 1b and Supplementary Table 2. Whereas above-chance-level performance was achieved for male patients with any anxiety disorder versus male HC classification (167 patients and 678 HC; 15 sites; average AUC = 0.63, $P = 0.007$), female patients with any anxiety disorder versus female HC classification (524 patients and 1,133 HC; 24 sites) failed to surpass chance level (average AUC = 0.57, $P = 0.218$). Supplementary LOSO-CV classifications also did not surpass chance level (Supplementary Table 3).

Psychotropic medication. The results of the transdiagnostic classifications based on current psychotropic medication use are summarized in Fig. 1c and Supplementary Table 2. The classification of patients who were unmedicated versus HC (641 patients and 1,969 HC; 32 sites) led to above-chance-level performance (average AUC = 0.59, $P = 0.013$), whereas classification of patients who were medicated versus HC (160 patients and 1,307 HC; 17 sites) failed to surpass chance level (average AUC = 0.59, $P = 0.22$). Similarly, the classification of patients who were medicated versus those who were unmedicated (251 medicated and 160 unmedicated patients; 17 sites) failed to surpass chance level (average AUC = 0.51, $P = 0.38$), as was the case for supplementary LOSO-CV classifications (Supplementary Table 3).

Severity. The results of the transdiagnostic classifications based on severity (State–Trait Anxiety Inventory–Trait (STAI-T) score of ≤48 for low-severity and >48 for high-severity anxiety) are summarized in Fig. 1d and Supplementary Table 2. Low-severity patient vs. HC classification (299 patients and 1,422 HC; 21 sites) performance was above chance level (average AUC = 0.59, $P = 0.016$), while high-severity patient vs. HC classification (272 patients and 1,449 HC; 22 sites) failed to surpass chance level (average AUC = 0.57, $P = 0.305$). Likewise, classification of low- vs. high-severity patients (299 low- and 270 high-severity patients; 21 sites) also failed to surpass chance level (average AUC = 0.52, $P = 0.235$), as was the case for the supplementary LOSO-CV classifications (Supplementary Table 3).

Comorbid depression. Finally, we performed case–control classifications separately for patients with and without comorbid major depression disorder (MDD). Neither of these classifications surpassed chance-level performance with either *K*-fold CV or LOSO-CV (see 'Comorbid depression subgroup analyses' in Supplementary Results for more details).

Feature importance

PD classifier. We investigated which brain regions contributed the most to above-chance-level classifications via sign-based consistency mapping. Four brain features emerged as most relevant for PD versus HC classification (Fig. 2 and Supplementary Table 4). Consistent positive signs of weights were found for CT in the left middle temporal gyrus and right rostral anterior cingulate cortex, which would imply a tendency of the support vector machine (SVM) for pushing the classification toward the PD class given increased CT in these regions.

https://doi.org/10.1038/s44220-023-00173-2

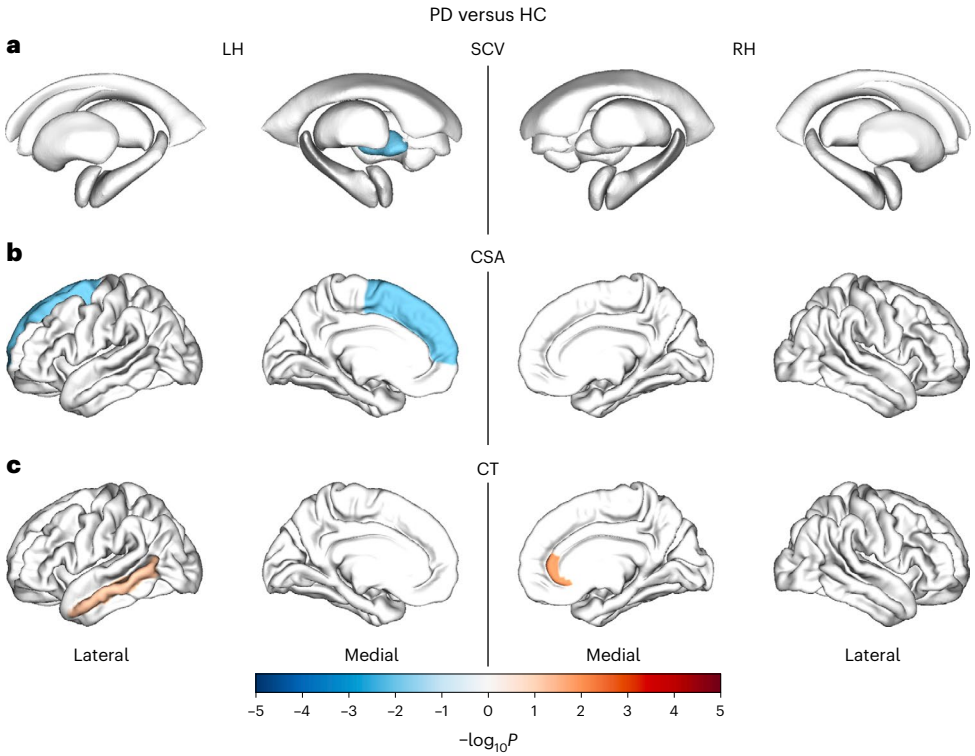

Fig. 2 | Brain regions contributing the most to classification of PD versus HC. Shown are $-\log_{10}P$ maps characterizing significant brain regions contributing the most to classification of PD versus HC, as determined by sign-based consistency mapping[79]. Hot colors indicate positive weights consistently assigned by the SVM that drive classification toward the patient class, and cold colors indicate negative weights that drive classification toward HC. **a**, Subcortical volumes. **b**, Cortical surface area. **c**, Cortical thickness. *Z*-statistic estimates from $\hat{p}$ values (importance scoring) were extracted following a *t*-test distribution and used to compute two-tailed $-\log_{10}P$. **a**–**c**, Full region names and associated statistics *Z*, $\hat{p}$ and $-\log_{10}P$ can be found in Supplementary Table 4. LH, left hemisphere; RH, right hemisphere. Figure created with the ENIGMA TOOLBOX (https://enigma-toolbox.readthedocs.io/en/latest/). Note that the derived sign-consistency *P* values do not correspond to a univariate group comparison and do not provide inherent localization information, but represent the most important features of the multivariate pattern used by the SVM classifier and should therefore be interpreted with caution (details in 'Feature importance' in Supplementary Methods).

Consistent negative signs of weights were found for the CSA of the left superior frontal gyrus and for SCV of the left pallidum, which would imply the opposite tendency of the SVM to assign participants to the HC class given increased SCV or CSA in these regions. Note that the derived sign-consistency *P* values do not correspond to a univariate group comparison, but represent the most important features of the multivariate pattern used by the SVM classifier, and should therefore be interpreted with caution (details in 'Feature importance' in Supplementary Methods).

Sex, medication and severity classifiers. The transdiagnostic classification of male patients versus male HC comprised seven significant features: positive signs of weights that pushed the classification toward the patient class were located in the CT of the right lateral occipital cortex and right superior frontal gyrus, and consistent negative signs of weights that pushed classification toward HC were found in both the CT and CSA of the right superior temporal gyrus, and in the CSA in the pars orbitalis, parahippocampal gyrus and cuneus cortex in the right hemisphere (Fig. 3 and Supplementary Table 5).

The transdiagnostic classification of unmedicated patients versus HC comprised 21 significant features (Fig. 4 and Supplementary Table 6). These included consistent positive signs of weights for the right caudate SCV; for CT in the right cuneus, left lingual and left superior temporal gyrus; and for CSA in the left middle temporal gyrus, left entorhinal, right superior parietal and right posterior-cingulate cortex (classification toward the patient class given increased CT or CSA). Negative signs of weights were found for the SCV of the left caudate and amygdala; for CT in the left inferior parietal cortex and right middle temporal gyrus; and for CSA in the right inferior temporal, supramarginal and left fusiform gyrus, bilateral temporal poles, right bank of the superior temporal sulcus, right paracentral lobule and left medial orbitofrontal and inferior parietal cortex (classification toward HC class given increased CT, CSA or SCV).

The transdiagnostic classification of patients with low-severity anxiety versus HC comprised six significant features (Figs. 5 and 6 and Supplementary Table 7). Positive signs of weights were found for CT in the right inferior parietal cortex and for CSA in the right rostral anterior cingulate and left entorhinal cortex (classification toward patient class given increased CT or CSA). Consistent negative signs of weights were found for CT and CSA in the right inferior temporal gyrus, and for CSA in the right pars triangularis (classification toward HC class given increased CT or CSA).

Discussion

In this study we benchmarked classification of youth with anxiety disorders based entirely on brain morphology, using structural MRI data from the large-scale multisite ENIGMA-Anxiety Working Group, comprising a heterogeneous sample of 3,343 participants from 32 international sites. Modest classification performance was achieved for PD versus HC (AUC = 0.62), but classification performances of SAD, GAD or any anxiety disorder (transdiagnostic) versus HC failed to surpass chance-level significance. However, modest transdiagnostic classifications were obtained in exploratory subgroup analyses of male patients versus male HC, unmedicated patients versus HC and

https://doi.org/10.1038/s44220-023-00173-2

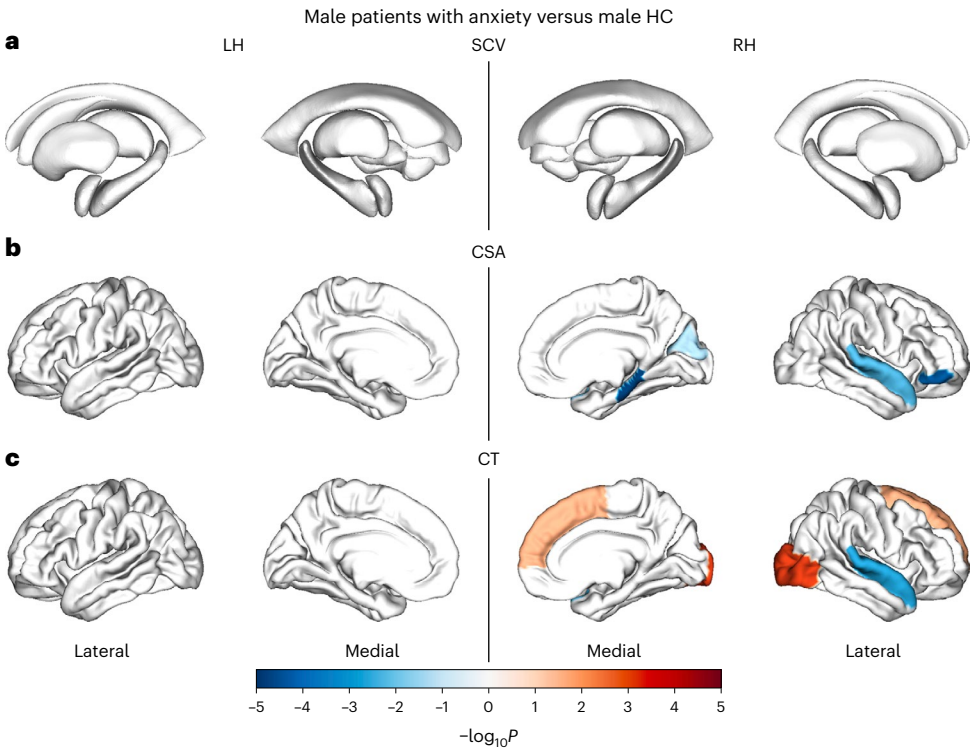

Fig. 3 | Brain regions contributing the most to classification of male patients versus male HC. Shown are $-\log_{10}P$ maps characterizing significant brain regions contributing the most to the classification of male patients versus male HC across working groups, as determined by sign-based consistency mapping[79]. Hot colors indicate positive weights consistently assigned by the SVM that drive classification toward the patient class, and cold colors indicate negative weights that drive classification toward HC. **a**, Subcortical volumes. **b**, Cortical surface area. **c**, Cortical thickness. *Z*-statistic estimates from $\hat{p}$ values (importance scoring) were extracted following a *t*-test distribution and used to compute two-tailed $-\log_{10}P$. **a**–**c**, Full region names and associated statistics *Z*, $\hat{p}$ and $-\log_{10}P$ can be found in Supplementary Table 5. LH, left hemisphere; RH, right hemisphere. Figure created with the ENIGMA TOOLBOX (https://enigma-toolbox.readthedocs.io/en/latest/). Note that the derived sign-consistency *P* values do not correspond to a univariate group comparison and do not provide inherent localization information, but represent the most important features of the multivariate pattern used by the SVM classifier and should therefore be interpreted with caution (details in 'Feature importance' in Supplementary Methods).

patients with low-severity anxiety versus HC (AUC = 0.59–0.63). These disorder-specific and transdiagnostic subgroup classifications were based on specific neuroanatomical features in fronto-striato-limbic and temporoparietal regions. This study shows that in a large, heterogeneous and multisite sample of youth with anxiety disorders, only modest classification performances can realistically be achieved based on neuroanatomical data. That said, the obtained classification accuracies do roughly translate into a medium effect size, with our results being on par with recent brain-based ML classifications of psychopathologies conducted in ENIGMA and other multisite consortia.

PD classifier

We identified several brain regions important for PD classification: CT of the left middle temporal gyrus and right rostral anterior cingulate cortex, along with the CSA of the left superior frontal gyrus and SCV of the left pallidum. The middle temporal gyrus is involved in several cognitive processes, including multimodal sensory integration, semantic processing and processing of emotions[32–34]. This structure is embedded within the temporal lobe and the broader limbic system involved in emotional responding, and is thought to play a role in PD pathophysiology[35,36]. Perturbed temporal lobe gray matter volume (GMV), specifically in the middle temporal gyrus and transverse temporal gyrus, are reported in adult patients with PD[36–38]. The anterior cingulate cortex is a key region for PD as it is implicated in the modulation of both normal and pathological anxiety, with voxel-based morphometry meta-analyses linking its disintegrity to PD diagnosis across age groups[37,39]. Several studies also report alterations in task-based anterior cingulate cortex activity in response to successful activation associated with treatment response toward cognitive–behavioral therapy[23,24]. The superior frontal gyrus is involved in cognitive control and emotion regulatory processes, and several studies link abnormal GMV in this region to PD pathophysiology[35,37,40]. Finally, the pallidum is part of the lenticular nucleus residing in the basal ganglia, and is a convergence point for limbic reward signals and is involved in diverse cognitive, affective and motor processes[41]. Voxel-based morphometry meta-analyses link GMV anomalies within the lenticular nucleus to clinical anxiety in general[39], and specifically in adult patients with PD[37]. Altogether, these findings suggest that brain regions deemed important for PD classification are plausible and in line with previous literature. However, considering the fairly modest classification performance of the model, caution should be exercised in interpreting these identified brain regions.

Sex, medication and severity classifiers

We investigated transdiagnostic classifications for subgroups with particular demographic or clinical characteristics (sex, medication use, symptom severity and comorbid MDD), with the goal of reducing heterogeneity across sites and to potentially boost performance. Transdiagnostic case–control classifications were significantly above chance level for some of the specified subgroups (AUC up to 0.63, $P < 0.05$) but not for the entire sample (AUC = 0.56, $P = 0.093$). As depicted in Figs. 2–5, the multivariate neurostructural pattern behind these transdiagnostic subgroup classifications comprised fronto-parieto-limbic regions previously linked to anxiety and its associated demographic (that is, sex) and clinical (that is, severity or medication use) characteristics[6,9,42–46].

https://doi.org/10.1038/s44220-023-00173-2

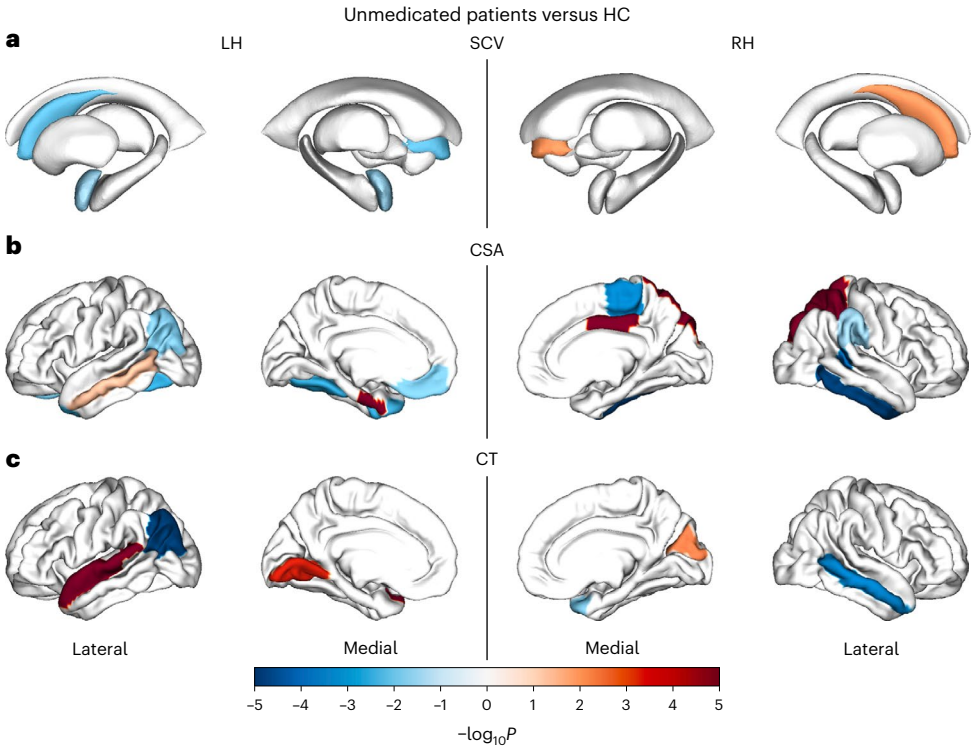

Fig. 4 | Brain regions contributing the most to classification of unmedicated patients versus HC. Shown are $-\log_{10}P$ maps characterizing significant brain regions contributing the most to classification of unmedicated patients versus HC across working groups, as determined by sign-based consistency mapping[79]. Hot colors indicate positive weights consistently assigned by the SVM that drive classification toward the patient class, and cold colors indicate negative weights that drive classification toward controls. **a**, Subcortical volumes. **b**, Cortical surface area. **c**, Cortical thickness. *Z*-statistic estimates from $\hat{p}$ values (importance scoring) were extracted following a *t*-test distribution and used to compute two-tailed $-\log_{10}P$. **a**–**c**, Full region names and associated statistics *Z*, $\hat{p}$ and $-\log_{10}P$ can be found in Supplementary Table 6. LH, left hemisphere; RH, right hemisphere. Figure created with the ENIGMA TOOLBOX (https://enigma-toolbox.readthedocs.io/en/latest/). Note that the derived sign-consistency *P* values do not correspond to a univariate group comparison and do not provide inherent localization information, but represent the most important features of the multivariate pattern used by the SVM classifier and should therefore be interpreted with caution (details in 'Feature importance' in Supplementary Methods).

Whereas transdiagnostic classification of male patients with anxiety versus male HC could be reliably achieved (AUC = 0.63, $P = 0.007$), this was not the case for female participants. Interestingly, such male-specific effects also emerged in a recent ENIGMA-Anxiety GAD case–control group comparison[42], while increased variability in female brain structure due to menarche, menstrual cycle and hormonal contraceptives is also suggested[47]. We also found that patients who were not medicated could be distinguished from HC (AUC = 0.59, $P = 0.013$), which was not the case for classification of patients on medication versus HC or patients who were medicated versus those who were unmedicated. One may speculate that heterogeneity in medication type, dose or duration among patients who are medicated could differentially influence brain structure and render classification more complicated. Alternatively, psychotropic medication use may normalize some neurostructural alterations in patients with anxiety[4,9]—a phenomenon also seen in other psychopathologies[48,49]—this may complicate classification of HC from their peers with unmedicated anxiety. Longitudinal studies that incorporate more detailed medication information may provide better insight into the short- and long-term effects of medication on brain structure in each specific disorder and across anxiety disorders.

Similarly, whereas patients with low-severity anxiety (as indexed by a below-median STAI-T score) could be distinguished from HC above chance level (AUC = 0.59, $P = 0.016$), this could not be achieved for classification of patients with high-severity anxiety versus HC or patients with high-severity versus low-severity anxiety. Although this result might be counterintuitive, it could be attributed to possible divergent brain abnormalities in patients with high-severity PD, SAD or GAD (for example, larger or smaller CT, CSA and/or SCV depending on the disorder), or to different variances in age ($P < 0.001$), sex ($P = 0.02$) and medication status ($P < 0.001$) we identified between patients with low- and high-severity anxiety. The heterogeneity in the high-severity group might be exacerbated by medication effects as severity and medication status were found to be positively correlated ($\varphi = 0.21$, $P < 0.001$), which could further limit the ability of the classifier to find common brain abnormalities across these patients. Nonetheless, the significance (or lack thereof, following permutation testing) for certain classifications could also reflect false positives or false negatives, and additional research is clearly needed to further explore, validate and refine these subgroup findings. Ideally, we would have also tested these subgroup classifications for each anxiety disorder separately, but this was simply not feasible given the small number of patients who would have remained after selection.

Finally, these subgroup classifications obtained with *K*-fold CV did not pass chance-level significance in our supplementary analyses using LOSO-CV. The discrepancy between the performances obtained with LOSO-CV and *K*-fold CV may be attributed to the different levels of generalizability being assessed. LOSO-CV reflects a realistic clinical scenario in which a model must generalize to entirely new samples[50]. In contrast, *K*-fold CV pools samples to identify common abnormalities that lead to optimal performance across all sites. Therefore, both CV approaches provide complementary sources of information. The lower performances obtained with LOSO-CV compared with *K*-fold CV were also found in similar multicenter classification studies on bipolar disorder (BP), autism, schizophrenia and obsessive–compulsive disorder (OCD)[28,29,50,51], and are likely due to increased heterogeneity between the training and test samples. By leaving out an entire site,

https://doi.org/10.1038/s44220-023-00173-2

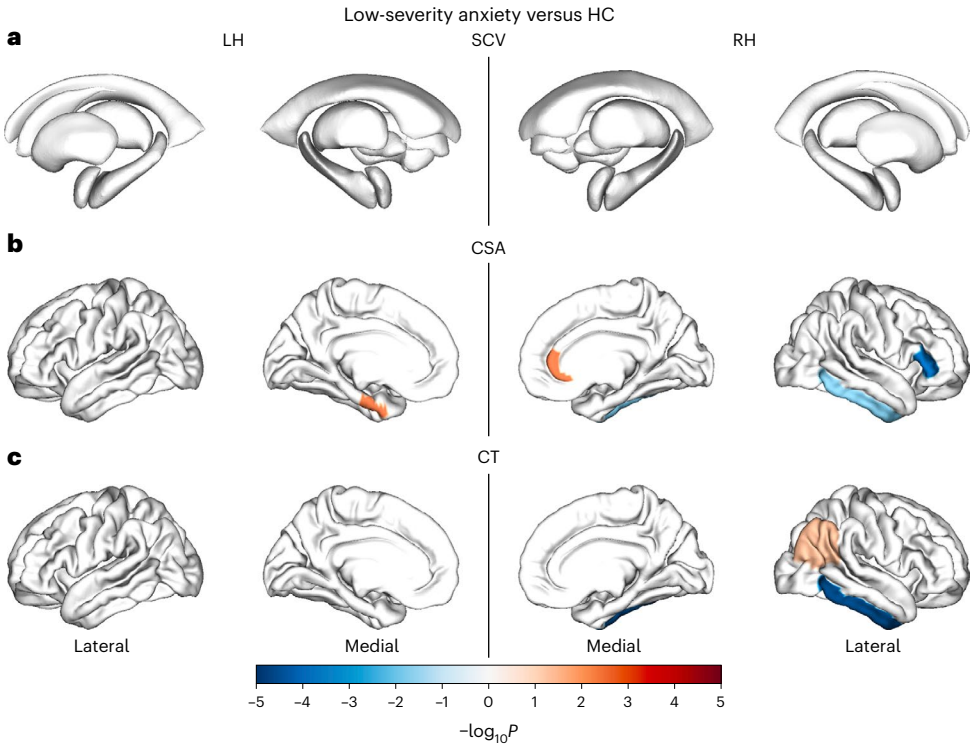

Fig. 5 | Brain regions contributing the most to classification of low-severity anxiety versus HC. Shown are $-\log_{10}P$ maps characterizing significant brain regions contributing the most to classification of low-severity anxiety versus HC across working groups, as determined by sign-based consistency mapping[79]. Hot colors indicate positive weights consistently assigned by the SVM that drive classification toward the patient class, and cold colors indicate negative weights that drive classification toward controls. **a**, Subcortical volumes. **b**, Cortical surface area. **c**, Cortical thickness. *Z*-statistic estimates from $\hat{p}$ values (importance scoring) were extracted following a *t*-test distribution and used to compute two-tailed $-\log_{10}P$. **a**–**c**, Full region names and associated statistics *Z*, $\hat{p}$ and $-\log_{10}P$ can be found in Supplementary Table 7. LH, left hemisphere; RH, right hemisphere. Figure created with the ENIGMA TOOLBOX (https://enigma-toolbox.readthedocs.io/en/latest/). Note that the derived sign-consistency *P* values do not correspond to a univariate group comparison and do not provide inherent localization information, but represent the most important features of the multivariate pattern used by the SVM classifier and should therefore be interpreted with caution (details in 'Feature importance' in Supplementary Methods).

specific demographic, clinical and scanner parameters from that site are not accounted for, whereas (site-stratified) *K*-fold CV ensures equal representation of each site in both training and test sets. In addition, it is worth noting that our study included many small sites. Consequently, LOSO-CV generated numerous small test folds that can lead to high variance in estimated performance, and may result in biased classification rates[52,53]. For these reasons, *K*-fold CV was incorporated in our main analyses, whereas LOSO-CV served as a supplementary method and complementary source of information. Nevertheless, given the discrepancies between the obtained performances with LOSO and *K*-fold CV for subgroup classifications, caution should be exercised in their interpretation.

Classifier performance and site effects

Our results are on par with other ENIGMA studies (on OCD, BP and MDD) that used FreeSurfer data for case–control classification in a multisite setting, with AUCs ranging between 0.62 and 0.72 (refs. 27–29). Similar to our results, none of the obtained performances reached the clinical threshold of AUC > 0.80 (ref. 54). These findings suggest that although there could be significant case–control differences in CT, CSA and/or SCV on the group level, parcellated structural MRI data may at the moment not fully allow for clinically relevant case–control distinctions at the individual level. Notably, the classification performances obtained here (AUCs up to 0.63) translate to effect sizes (Cohen's $d = 0.47$; medium effect size; see ref. 30 for AUC to effect size conversion) typically larger than those obtained in previous ENIGMA univariate case–control analyses of FreeSurfer data among psychiatric populations. This includes case–control analyses of OCD (maximum Cohen's $d = -0.33$ for CSA), BP (maximum Cohen's $d = -0.29$ for SCV), schizophrenia (maximum Cohen's $d = -0.46$ for SCV), MDD (maximum Cohen's $d = -0.2$ for SCV, but up to −0.57 for CSA in adolescents versus matched controls) and attention deficit hyperactivity disorder (maximum Cohen's $d = -0.19$ for SCV)[26,55–59]. However, one should be cautious in directly comparing the above-mentioned effect sizes, as our AUC-converted effect size was based on a multivariate method estimated through CV (out of sample), whereas the case–control effect sizes are in sample and univariate in nature. Future studies could apply more fine-grained neuroanatomical features (for example, voxel-wise or vertex-wise maps), measures of brain connectivity and network function or even multimodal data in combination with more sophisticated classification methods (for example, deep neural networks)[15,27]. A combination of these options is postulated to improve classification performances compared with shallow ML algorithms applied to low-resolution data[15].

Previous structural and functional MRI-based classification studies on anxiety disorders have claimed to distinguish patients from HC, with AUCs reaching over 0.80 (refs. 17–24). However, these studies relied on small samples ($N < 100$) susceptible to unstable performance estimations during CV[53,60], in which inflated performances are likely to be over-represented through publication bias or insufficient testing[61–63]. In addition, studies typically used data from a single research site, and performance typically drops when models are tested on unseen data from other sites. The large-scale, multisite ENIGMA database used here provides clinically representative information, which may be used by ML algorithms to identify multivariate patterns generalizable to patients across different sites[27–29]. Having a large sample

https://doi.org/10.1038/s44220-023-00173-2

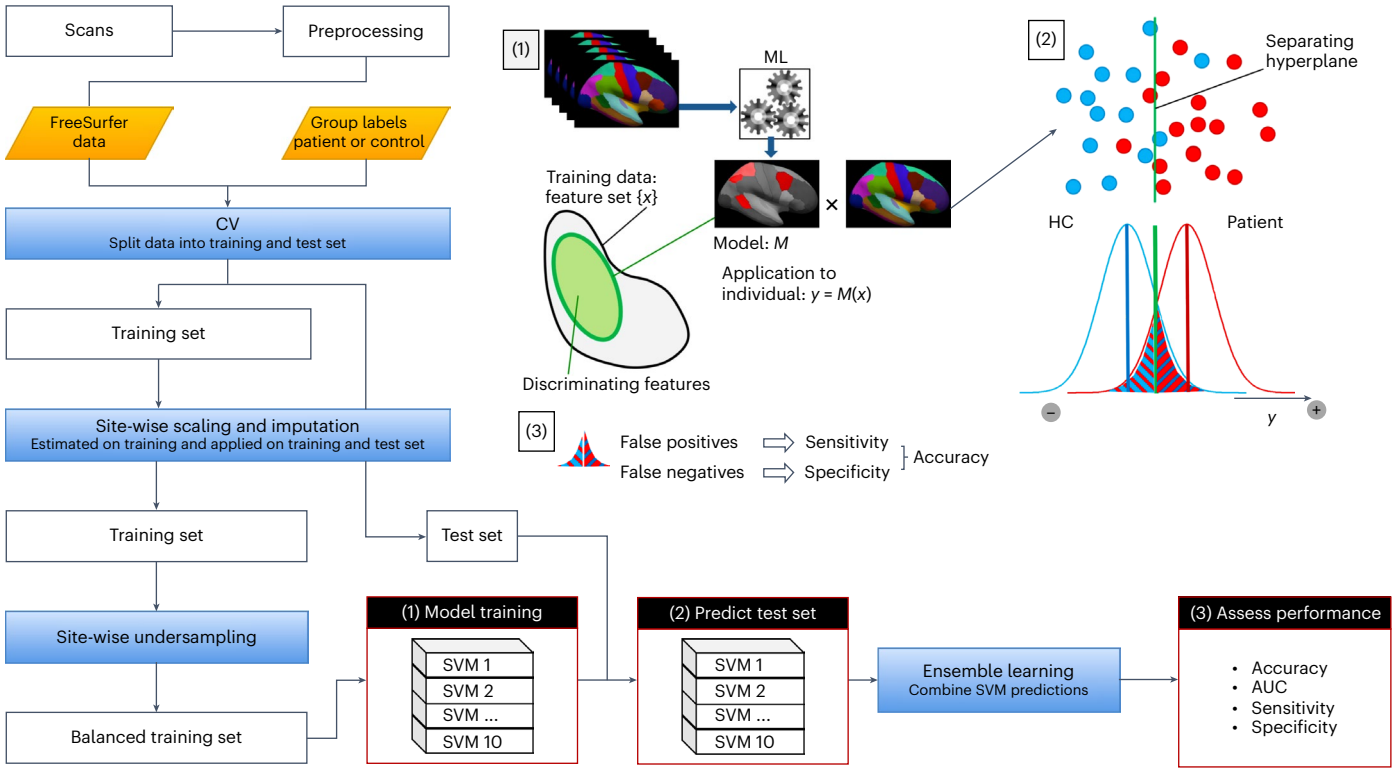

Fig. 6 | Simplified visual representation of the ML pipeline. (1) ML algorithm training based on labeled FreeSurfer features (CT or CSA and SCV); (2) this results in a model *M* that uses a discriminative subset of these features (feature vector) to classify patients versus controls[30] via an optimal separating hyperplane within the entire feature space[30]. Output value *y* (distance to hyperplane) indicates how precisely or accurately the model performs when applied to a participant's feature vector (blue represents HC; red represents patients)[30]. (3) False-negative and false-positive classifications are enclosed within overlapping parts of the separating hyperplane distribution[30]. Figure adapted with permission from ref. 30 under a Creative Commons license CC BY 4.0.

size and thus more data for model training and testing is typically beneficial for ML, and leads to more reliable performance estimates[50,51,64]. However, pooling existing data in a multisite context like ENIGMA may also reveal that classification methods have poorer true performance in reality, due to the realistic heterogeneity in sample characteristics, hardware and methodology that is actually represented[27,28,30]. These global between-site differences, referred to as 'site effects', can hamper the ability of the classifier to find common brain abnormalities between patients and HC across different sites in the performed classification tasks (that is, disorder-specific and transdiagnostic classifications). Site effects can also leave confounds in the data that obscure interpretations, impair the generalizability and reproducibility of classification models and lead to biased performance[27,65].

We addressed site effects by standardizing neuroimaging data from each site according to their HC reference group. Without site-wise scaling, the SVM was able to accurately predict to which site a given participant belongs, but not when site-wise scaling was applied (Supplementary Table 8). Performances obtained for our main classifications using *K*-fold CV were comparable to LOSO-CV, further indicating that our standardization approach effectively harmonized data across sites; but this was not the case for our exploratory subgroup analyses in which none of the classification performances obtained with LOSO-CV surpassed chance level. However, we also accounted for site effects in our permutation-testing framework by restricting the exchangeability of class labels to each site so that remaining site effects would be incorporated in the obtained null distribution (see 'Site effects' in Supplementary Discussion)[66]. There are other methods available to handle site effects, such as ComBat and normative modeling, but the assumptions for these techniques (for example, $n > 25$ subjects per site, balanced samples and overlapping distribution of covariates across sites) are violated here, rendering them unsuitable for this study[65,67]. Future classification studies on data from the ENIGMA-Anxiety Working Group using larger samples could investigate the feasibility of other site-harmonization techniques.

Strengths and limitations

The key advantages of this study include the large worldwide sample ($N = 3{,}343$ from 32 sites), access to individual-level data and the conservative nature of the analyses wherein possible site effects, data leakage and overfitting were stringently confronted. The international heterogeneity and multisite nature allowed us to test ML classifications on a wide range of participants, more closely resembling the real-world clinical situation. Notwithstanding these strengths, there are several limitations to consider. We used a global and retrospectively pooled data sample, where no harmonization of scanning acquisition or sample inclusion was performed. These sources of heterogeneity may have limiting effects on classification performance, but also provide an opportunity for a more realistic evaluation across samples that better represent the wide range of characteristics seen in the real-world population. This would be considered a strength, as it is necessary to evaluate whether results are reliable across a variety of hardware and protocols.

As we specifically focused on youth (10–25 years), we had to exclude participants from each sample outside this age range and ended up with a relatively large number of sites with small sample sizes. This limited our options for site-effect harmonization, and led to insufficient sample sizes per site to investigate subgroup classifications for each anxiety disorder separately. Also, whereas developmentally driven heterogeneity may have been at play, the sample size (especially patients) made it practically impossible to formally test this, as we

lacked sufficient data for training or testing folds per meaningful age grouping. However, supplementary analyses were run to assess the link between age and classification performance within the transdiagnostic sample among different age groups (early adolescence = 10–14; middle adolescence = 15–17; late adolescence and young adulthood = 18–25)[31,68,69]. As seen in Supplementary Fig. 2, there was no statistically significant association between the obtained performances and age groups ($r = -0.50$, $P = 0.667$), suggesting that the model performed equally well across the age spectrum investigated. Also, we only had access to regional brain measures (FreeSurfer features), not raw or voxel-wise data or other brain-imaging modalities (for example, function or connectivity). Incorporation of these additional data forms, and use of more sophisticated ML algorithms (for example, deep learning) better suited to handling massively multimodal data, might further improve classification performances reported here[15]. In addition, we would have ideally also run subgroup classifications for each anxiety disorder separately to gain deeper insights into disorder-specific heterogeneity, as a function of clinical or sociodemographic characteristics. However, meaningful and consistent subgroup classifications for each anxiety disorder proved infeasible given the small number of patients who would have remained after selection (Supplementary Table 9). This precluded us from effectively probing disorder-specific heterogeneity in each anxiety disorder, something that future work should consider.

Finally, whereas brain data were available for each participant, this was not the case for detailed clinical or sociodemographic data, making it difficult to test the added value of brain-based classifiers compared with conventional clinical or sociodemographic information. This is especially relevant for ML classifications of clinically anxious youth, as these types of data have been shown to produce comparable performances[44]. Although the purely brain-based classifications as reported here may serve as a benchmark, the potential added value of routine clinical data (predictive, inexpensive and easy to collect compared with neuroimaging data) should be acknowledged and ideally incorporated in future classification studies[44]. Recent studies in psychiatric youth have shown that using a multimodal approach that combines MRI with other data modalities (for example, questionnaires, cognitive tests and genetics) can lead to better diagnostic classifications[44]. Moreover, the only data-driven, computational diagnostic tool currently approved by the US Food and Drug Administration for child and adolescent psychiatry is one that combines neuro-bio-behavioral data[44,70]. Future directions for clinical utility could also involve assessments that are specifically tailored to youth (for example, smart phones, social media, virtual reality and gaming) along with net-benefit analyses[44]. An exciting approach within the field is one that entails multiple sequential steps informed by big data analyses with ML algorithms, which implement cost-effective sequences of clinical and neuro-bio-behavioral assessments to maximize predictive power and clinical utility[44,71]. Further research in this area could help move the field toward true translational impact.

Conclusion

This study shows that in a large, heterogeneous and multisite sample of youth with anxiety disorders, only modest classification performances can realistically be achieved with machine learning using neuroanatomical data. That said, the obtained classification accuracies do roughly translate into a medium effect size (Cohen's $d = 0.47$; see ref. 30 for AUC to effect size conversion), with our results being on par with recent brain-based ML classifications of psychopathologies conducted in ENIGMA and other multisite consortia[27–29]. This classification study sets a baseline for the ENIGMA-Anxiety Working Group, and is an important step toward the development of models that could ultimately inform early detection, prevention and care among clinically anxious youth. Future work should probe and evaluate the added value compared with more conventional diagnostic tools, such as structured clinical interviews and questionnaires.

Methods

Study sample

Participants were included from three ENIGMA-Anxiety Working Groups: PD (12 sites), 112 patients and 813 controls; GAD (16 sites), 465 patients and 1,084 controls; and SAD (13 sites), 259 patients and 610 controls (the STROBE statement flowchart is provided in Supplementary Information). All participants included in the analysis were between 10 and 25 years of age. HC were free of past and present psychopathology and psychotropic medication use at the time of inclusion. Comorbid anxiety disorders (PD, GAD or SAD) were present in some patients, and their assigned primary diagnosis (anxiety disorder according to the *Diagnostic and Statistical Manual of Mental Disorders, Fourth or Fifth Edition*)[72,73] corresponds to their respective working group. Participants' demographic and clinical characteristics are summarized in Table 1. For an overview per site, see Supplementary Table 1. Data on ethnicity were not collected by the majority of sites and therefore could not be robustly analyzed, although the global nature of our dataset seems to afford adequate ethnic representation. This study was conducted in accordance with the Declaration of Helsinki. Individual studies were previously approved by the relevant local ethical review boards, with this study falling under the guidelines of those ethical review boards for secondary use of collected data. All participants or caretakers provided written informed consent.

Image acquisition and processing

Structural T1-weighted three-dimensional brain MRI scans were acquired and processed either locally at each site (PD and SAD data) or centrally (GAD data) using standardized protocols for harmonized analysis and quality control. Images were acquired at different field strengths (1.5 Tesla or 3.0 Tesla); sample-specific acquisition parameters are listed elsewhere[10]. Regional mean CT, CSA and SCV were extracted from the brain images using FreeSurfer (v.5.3 or v.6.0)[74]. Parcellations or segmentations were visually inspected and statistically evaluated for outliers. For each subject, CSA and CT were calculated for 68 cortical Desikan–Killiany atlas-based regions (34 per hemisphere)[75]. In addition, gray matter SCVs were extracted for seven structures per hemisphere (nucleus accumbens, putamen, pallidum, caudate, thalamus, amygdala and hippocampus), along with lateral ventricle volumes. This yielded a total of 152 FreeSurfer features per subject for classification.

Multivariate brain-based classification

Primary analyses. As shown in Fig. 6, fully brain-based classification tasks were performed with linear SVMs, one of the most commonly and successfully used algorithms in the field[15,44,76]. We performed disorder-specific case–control classification for each working group separately (PD versus HC, GAD versus HC and SAD versus HC). We also performed a transdiagnostic classification for patients with any anxiety disorder (PD or GAD or SAD) versus HC using pooled samples across the three working groups. A total of 561 participants (545 HC, 5 patients with GAD and 11 patients with SAD) whose data were present in more than one working group were included only once in transdiagnostic classifications to avoid the use of duplicated data. Participants with missing data that had less than 75% of combined cortical and subcortical FreeSurfer features were excluded (14 HC and 3 patients with PD).

Classification performance was evaluated using stratified five-times-repeated five-fold CV, following guidelines for the field[52,77]. The proportion of patients and controls from each site was (approximately) maintained for each fold. We also evaluated to what extent the aforementioned classifiers were able to generalize to unseen sites using LOSO-CV, using this supplementary validation strategy to provide complementary information to K-fold CV (Supplementary Results). In each LOSO-CV fold, one site is left out for model testing and the remaining sites are used for training, and this is repeated so that each site is used as a test set once. Classification performance was measured

using the AUC, balanced accuracy, sensitivity and specificity, which were calculated for each CV iteration on the testing set and averaged across CV iterations.

Features were standardized using the training set in a site-wise manner by calculating the mean and s.d. of each feature on the HC from each site separately, and then applied to standardize test and training data for both patients and HC of the same site to obtain harmonized and comparably interpretable features. Sites that collected MRI data across different scanners were handled as individual sites in our analyses. Standardization was thus performed in a normative manner, where brain measures from HC of each site are used as a reference point against which patients can be compared. This was done to account for site effects (for example, differences in data acquisition protocols and inclusion criteria) that could affect classification performance[27,65]. There are alternative methods to handle site effects, such as ComBat and normative modeling[65]; however, the assumptions for these techniques are violated here, rendering them unsuitable for this study (this is further addressed in 'Discussion'). Only participants from sites that had data for both HC and patients, and at least ten HC, were included for each classification task to ensure sufficient data for standardization. Missing features were mean imputed using the full training set. Site-wise undersampling was applied on the training set to account for class imbalance for each site separately so that an equal number of samples was used from both classes. The undersampling procedure was repeated ten times for each training set within the CV procedure, resulting in ten SVM models trained using different (balanced) training sets and evaluated on the same test set (no undersampling). Classifications across the resulting SVM models were combined using an ensemble approach by taking the median across the decision values obtained for the predictions of the test set. We applied the SVM classification with the regularization parameter (C) set to 1, following general recommendations from the field[52]. SVM class weights for C were set to 'balanced' mode to automatically adjust weights inversely proportional to class frequencies in the input data. More details on the classification procedure are provided in 'Full dataset case–control analysis' in Supplementary Methods. Label permutation testing with 1,000 iterations[78] was used to test whether classification performance (AUC) was significantly above chance level ($\alpha = 0.05$), with our hypothesis being above chance-level classification of patients versus HC. To assess which brain regions contributed most to the classification model, we applied sign-based consistency mapping as per ref. 79 (see details in 'Feature importance' in Supplementary Methods). We only report sign-based feature importance for classifications that passed label permutation testing.

Exploratory subgroup analyses. To explore the effects of demographic and clinical heterogeneity on brain-based classification performance, we performed subgroup analyses on sex, current psychotropic medication use, anxiety severity at time of scanning (STAI-T[80]; median split on patients' scores (median = 48) produced high- or low-severity groups) and current or lifetime comorbid MDD. Subgroup analyses were restricted to the transdiagnostic sample (any anxiety disorder versus HC across working groups) as limited data would remain when investigating these subgroups in each disorder group separately. The classification procedure itself (that is, site-wise scaling, imputation, undersampling and ensemble learning) was as described above. Classifications included female patients versus female HC, male patients versus male HC, patients who were medicated versus HC, patients who were unmedicated versus HC, patients who were unmedicated versus those who were medicated, patients with low-severity anxiety (STAI-T ≤ 48) versus HC, patients with high-severity anxiety (STAI-T > 48) vs. HC, patients with low-severity anxiety versus those with high-severity anxiety, patients with comorbid MDD versus HC and patients without comorbid MDD versus HC.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data used in this study are in principle publicly available within the ENIGMA Consortium. For this study, all site-level data were first transferred to a secure data environment (Amsterdam UMC) and subsequently subjected to the final mega-analyses, all with permission from individual sites. However, some data-sharing restrictions are in place. These include restrictions imposed by (1) site-specific consent documentation, ethical review boards and institutional processes, (2) along with national/international data-sharing legislations (for example, General Data Protection Regulation). Some of these restrictions may require a signed material transfer agreement for limited and predefined data use. However, data sharing might still be possible, requiring submission of a detailed analysis plan to the ENIGMA-Anxiety Consortium. If approved, access to relevant data is provided, depending on data availability, local principal investigator approval and compliance with all supervening regulations. Requests or questions regarding data availability or sharing should be sent to the corresponding author.

Code availability

Relevant analysis codes can be found at https://github.com/WillemB2104.

References

1. Zacharek, S. J., Kribakaran, S., Kitt, E. R. & Gee, D. G. Leveraging big data to map neurodevelopmental trajectories in pediatric anxiety. *Dev. Cogn Neurosci.* **50**, 100974 (2021).
2. Merikangas, K. R. et al. Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *J. Am. Acad. Child Adolesc. Psychiatry* **49**, 980–989 (2010).
3. Vigo, D., Thornicroft, G. & Atun, R. Estimating the true global burden of mental illness. *Lancet Psychiatry* **3**, 171–178 (2016).
4. Strawn, J. R., Lu, L., Peris, T. S., Levine, A. & Walkup, J. T. Research review: pediatric anxiety disorders – what have we learnt in the last 10 years? *J. Child Psychol. Psychiatry* **62**, 114–139 (2021).
5. Hafstad, G. S. & Augusti, E. M. A lost generation? COVID-19 and adolescent mental health. *Lancet Psychiatry* **8**, 640–641 (2021).
6. Strawn, J. R. et al. Neurobiology of pediatric anxiety disorders. *Curr. Behav. Neurosci. Rep.* **1**, 154–160 (2014).
7. Cosgrove, V. E. et al. Structure and etiology of co-occurring internalizing and externalizing disorders in adolescents. *J. Abnorm. Child Psychol.* **39**, 109–123 (2011).
8. Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G. & Angold, A. Prevalence and development of psychiatric disorders in childhood and adolescence. *Arch. Gen. Psychiatry.* **60**, 837–844 (2003).
9. Craske, M. G. et al. Anxiety disorders. *Nat. Rev. Dis. Primers* **3**, 17024 (2017).
10. Bas-Hoogendam, J. M. et al. ENIGMA-Anxiety working group: rationale for and organization of large-scale neuroimaging studies of anxiety disorders. *Hum. Brain Mapp.* **43**, 83–112 (2022).
11. Janssen, R. J., Mourao-Miranda, J. & Schnack, H. G. Making individual prognoses in psychiatry using neuroimaging and machine learning. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **3**, 798–808 (2018).
12. Marquand, A. F. et al. Conceptualizing mental disorders as deviations from normative functioning. *Mol. Psychiatry* **24**, 1415–1424 (2019).
13. Marquand, A. F., Rezek, I., Buitelaar, J. & Beckmann, C. F. Understanding heterogeneity in clinical cohorts using normative models: beyond case–control studies. *Biol. Psychiatry* **80**, 552–561 (2016).

14. Paulus, M. P. & Thompson, W. K. The challenges and opportunities of small effects: the new normal in academic psychiatry. *JAMA Psychiatry* **76**, 353–354 (2019).
15. Bzdok, D. & Meyer-Lindenberg, A. Machine learning for precision psychiatry: opportunities and challenges. *Biol. Psychiatry Cogn Neurosci. Neuroimaging* **3**, 223–230 (2018).
16. Orru, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G. & Mechelli, A. Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci. Biobehav. Rev.* **36**, 1140–1152 (2012).
17. Zhang, W. et al. Diagnostic prediction for social anxiety disorder via multivariate pattern analysis of the regional homogeneity. *Biomed. Res. Int.* **2015**, 763965 (2015).
18. Liu, F. et al. Multivariate classification of social anxiety disorder using whole brain functional connectivity. *Brain Struct. Funct.* **220**, 101–115 (2015).
19. Xing, M., Fitzgerald, J. M. & Klumpp, H. Classification of social anxiety disorder with support vector machine analysis using neural correlates of social signals of threat. *Front. Psychiatry* **11**, 144 (2020).
20. Frick, A. et al. Classifying social anxiety disorder using multivoxel pattern analyses of brain function and structure. *Behav. Brain Res.* **259**, 330–335 (2014).
21. Hilbert, K., Lueken, U., Muehlhan, M. & Beesdo-Baum, K. Separating generalized anxiety disorder from major depression using clinical, hormonal, and structural MRI data: a multimodal machine learning study. *Brain Behav.* **7**, e00633 (2017).
22. Pantazatos, S. P., Talati, A., Schneier, F. R. & Hirsch, J. Reduced anterior temporal and hippocampal functional connectivity during face processing discriminates individuals with social anxiety disorder from healthy controls and panic disorder, and increases following treatment. *Neuropsychopharmacology.* **39**, 425–434 (2014).
23. Lueken, U. et al. Neurobiological markers predicting treatment response in anxiety disorders: a systematic review and implications for clinical application. *Neurosci. Biobehav. Rev.* **66**, 143–162 (2016).
24. Lueken, U., Hilbert, K., Wittchen, H. U., Reif, A. & Hahn, T. Diagnostic classification of specific phobia subtypes using structural MRI data: a machine-learning approach. *J. Neural Transm.* **122**, 123–134 (2015).
25. Thompson, P. M. et al. The ENIGMA consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav.* **8**, 153–182 (2014).
26. Thompson, P. M. et al. ENIGMA and global neuroscience: a decade of large-scale studies of the brain in health and disease across more than 40 countries. *Transl. Psychiatry* **10**, 100 (2020).
27. Belov, V. et al. Global multi-site benchmark classification of major depressive disorder using machine learning on cortical and subcortical features of 5,365 participants from the ENIGMA MDD dataset. Preprint at arxiv.org/abs/2206.08122 (2022).
28. Bruin, W. B. et al. Structural neuroimaging biomarkers for obsessive-compulsive disorder in the ENIGMA-OCD consortium: medication matters. *Transl. Psychiatry* **10**, 342 (2020).
29. Nunes, A. et al. Using structural MRI to identify bipolar disorders—13 site machine learning study in 3020 individuals from the ENIGMA Bipolar Disorders Working Group. *Mol. Psychiatry* **25**, 2130–2143 (2020).
30. Schnack, H. G. & Kahn, R. S. Detecting neuroimaging biomarkers for psychiatric disorders: sample size matters. *Front. Psychiatry* **7**, 50 (2016).
31. Sawyer, S. M., Azzopardi, P. S., Wickremarathne, D. & Patton, G. C. The age of adolescence. *Lancet Child Adolesc. Health* **2**, 223–228 (2018).
32. Wang, S. et al. Emotional intelligence mediates the association between middle temporal gyrus gray matter volume and social anxiety in late adolescence. *Eur. Child Adolesc. Psychiatry* **30**, 1857–1869 (2021).
33. Grabe, H. J. et al. Alexithymia and brain gray matter volumes in a general population sample. *Hum. Brain Mapp.* **35**, 5932–5945 (2014).
34. Ren, J. et al. The function of the hippocampus and middle temporal gyrus in forming new associations and concepts during the processing of novelty and usefulness features in creative designs. *NeuroImage* **214**, 116751 (2020).
35. Graeff, F. G. & Del-Ben, C. M. Neurobiology of panic disorder: from animal models to brain neuroimaging. *Neurosci. Biobehav. Rev.* **32**, 1326–1335 (2008).
36. Sobanski, T. et al. Temporal and right frontal lobe alterations in panic disorder: a quantitative volumetric and voxel-based morphometric MRI study. *Psychol. Med.* **40**, 1879–1886 (2010).
37. Wang, X. et al. Distinct grey matter volume alterations in adult patients with panic disorder and social anxiety disorder: a systematic review and voxel-based morphometry meta-analysis. *J. Affect. Disord.* **281**, 805–823 (2021).
38. de Carvalho, M. R. et al. Current findings of fMRI in panic disorder: contributions for the fear neurocircuitry and CBT effects. *Expert Rev. Neurother.* **10**, 291–303 (2010).
39. Radua, J., van den Heuvel, O. A., Surguladze, S. & Mataix-Cols, D. Meta-analytical comparison of voxel-based morphometry studies in obsessive-compulsive disorder vs other anxiety disorders. *Arch. Gen. Psychiatry* **67**, 701–711 (2010).
40. Protopopescu, X. et al. Increased brainstem volume in panic disorder: a voxel-based morphometric study. *NeuroReport* **17**, 361–363 (2006).
41. Smith, K. S., Tindell, A. J., Aldridge, J. W. & Berridge, K. C. Ventral pallidum roles in reward and motivation. *Behav. Brain Res.* **196**, 155–167 (2009).
42. Harrewijn, A. et al. Cortical and subcortical brain structure in generalized anxiety disorder: findings from 28 research sites in the ENIGMA-Anxiety Working Group. *Transl. Psychiatry* **11**, 502 (2021).
43. Duval, E. R., Javanbakht, A. & Liberzon, I. Neural circuits in anxiety and stress disorders: a focused review. *Ther. Clin. Risk Manag.* **11**, 115–126 (2015).
44. Dwyer, D. & Koutsouleris, N. Annual research review: translational machine learning for child and adolescent psychiatry. *J. Child Psychol. Psychiatry* **63**, 421–443 (2022).
45. Grupe, D. W. & Nitschke, J. B. Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. *Nat. Rev. Neurosci.* **14**, 488–501 (2013).
46. Xie, S., Zhang, X., Cheng, W. & Yang, Z. Adolescent anxiety disorders and the developing brain: comparing neuroimaging findings in adolescents and adults. *Gen. Psychiatr.* **34**, e100411 (2021).
47. Rehbein, E., Hornung, J., Sundstrom Poromaa, I. & Derntl, B. Shaping of the female human brain by sex hormones: a review. *Neuroendocrinology* **111**, 183–206 (2021).
48. Dusi, N., Barlati, S., Vita, A. & Brambilla, P. Brain structural effects of antidepressant treatment in major depression. *Curr. Neuropharmacol.* **13**, 458–465 (2015).
49. Hajek, T. & Weiner, M. W. Neuroprotective effects of lithium in human brain? Food for thought. *Curr. Alzheimer Res.* **13**, 862–872 (2016).
50. Abraham, A. et al. Deriving reproducible biomarkers from multi-site resting-state data: an Autism-based example. *NeuroImage* **147**, 736–745 (2017).
51. Nieuwenhuis, M. et al. Classification of schizophrenia patients and healthy controls from structural MRI scans in two large independent samples. *NeuroImage* **61**, 606–612 (2012).

52. Varoquaux, G. et al. Assessing and tuning brain decoders: cross-validation, caveats, and guidelines. *NeuroImage* **145**, 166–179 (2017).
53. Flint, C. et al. Systematic misestimation of machine learning performance in neuroimaging studies of depression. *Neuropsychopharmacology.* **46**, 1510–1517 (2021).
54. Botteron, K. et al. *Consensus Report of the APA Work Group on Neuroimaging Markers of Psychiatric Disorders* (American Psychiatric Association, 2012).
55. Boedhoe, P. S. W. et al. Cortical abnormalities associated with pediatric and adult obsessive–compulsive disorder: findings from the ENIGMA Obsessive–Compulsive Disorder Working Group. *Am. J. Psychiatry.* **175**, 453–462 (2018).
56. Hibar, D. P. et al. Cortical abnormalities in bipolar disorder: an MRI analysis of 6503 individuals from the ENIGMA Bipolar Disorder Working Group. *Mol. Psychiatry* **23**, 932–942 (2018).
57. van Erp, T. G. et al. Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Mol. Psychiatry* **21**, 547–553 (2016).
58. Schmaal, L. et al. Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Mol. Psychiatry* **22**, 900–909 (2017).
59. Hoogman, M. et al. Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: a cross-sectional mega-analysis. *Lancet Psychiatry* **4**, 310–319 (2017).
60. Varoquaux, G. Cross-validation failure: small sample sizes lead to large error bars. *NeuroImage* **180**, 68–77 (2018).
61. Koppe, G., Meyer-Lindenberg, A. & Durstewitz, D. Deep learning for small and big data in psychiatry. *Neuropsychopharmacology.* **46**, 176–190 (2021).
62. Vieira, S., Liang, X., Guiomar, R. & Mechelli, A. Can we predict who will benefit from cognitive-behavioural therapy? A systematic review and meta-analysis of machine learning studies. *Clin. Psychol. Rev.* **97**, 102193 (2022).
63. Eitel, F., Schulz, M. A., Seiler, M., Walter, H. & Ritter, K. Promises and pitfalls of deep neural networks in neuroimaging-based psychiatric research. *Exp. Neurol.* **339**, 113608 (2021).
64. Woo, C. W., Chang, L. J., Lindquist, M. A. & Wager, T. D. Building better biomarkers: brain models in translational neuroimaging. *Nat. Neurosci.* **20**, 365–377 (2017).
65. Bayer, J. M. M. et al. Site effects how-to and when: an overview of retrospective techniques to accommodate site effects in multi-site neuroimaging analyses. *Front. Neurol.* **13**, 923988 (2022).
66. Winkler, A. M., Ridgway, G. R., Webster, M. A., Smith, S. M. & Nichols, T. E. Permutation inference for the general linear model. *NeuroImage* **92**, 381–397 (2014).
67. Zindler, T., Frieling, H., Neyazi, A., Bleich, S. & Friedel, E. Simulating ComBat: how batch correction can lead to the systematic introduction of false positive results in DNA methylation microarray studies. *BMC Bioinf.* **21**, 271 (2020).
68. Steinberg, L. D. *Adolescence* (McGraw-Hill, 1993).
69. Patton, G. C. et al. Adolescence and the next generation. *Nature* **554**, 458–466 (2018).
70. Abbas, H., Garberson, F., Liu-Mayo, S., Glover, E. & Wall, D. P. Multi-modular AI approach to streamline autism diagnosis in young children. *Sci. Rep.* **10**, 5014 (2020).
71. Koutsouleris, N. et al. Multimodal machine learning workflows for prediction of psychosis in patients with clinical high-risk syndromes and recent-onset depression. *JAMA Psychiatry* **78**, 195–209 (2021).
72. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (American Psychiatric Association, 2000).
73. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (American Psychiatric Association, 2013).
74. Fischl, B. et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341–355 (2002).
75. Desikan, R. S. et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980 (2006).
76. Claude, L. A., Houenou, J., Duchesnay, E. & Favre, P. Will machine learning applied to neuroimaging in bipolar disorder help the clinician? A critical review and methodological suggestions. *Bipolar Disord.* **22**, 334–355 (2020).
77. Poldrack, R. A., Huckins, G. & Varoquaux, G. Establishment of best practices for evidence for prediction: a review. *JAMA Psychiatry* **77**, 534–540 (2020).
78. Ojala, M. & Garriga, G. C. Permutation tests for studying classifier performance. *J. Mach. Learn. Res.* **11**, 1833–1863 (2010).
79. Gomez-Verdejo, V., Parrado-Hernandez, E., Tohka, J. & Alzheimera's Disease Neuroimaging Initiative. Sign-consistency based variable importance for machine learning in brain imaging. Neuroinformatics 17, 593–609 (2019).
80. Spielberger, C. D. *State–Trait Anxiety Inventory Bibliography*, 2nd ed. (Consulting Psychologists Press, 1989).

Acknowledgements

This study was financially supported by The Netherlands Organization for Health Research and Development—ZonMw (research fellowship number 06360322210035 to M. Aghajani), Dutch Research Council NWO (SSH Open Competition number 15810 to M. Aghajani), Leiden University Fund (project Youth Mental Health Meets Big Data Analytics, grant number LUF23075-5-306 to M. Aghajani), Leiden University Fund (project grant number W213085-5 to M. Aghajani), Amsterdam Neuroscience (Amsterdam Neuroscience Alliance Grant Project to M. Aghajani and G.A.v.W.), the Dutch Research Council NWO (Rubicon grant number 019.201SG.022 to J.M.B.-H. and 452020227 to L.K.M.H.), the Carnegie Corporation of New York (to N.A.G.; the statements made and views expressed are solely the responsibility of the author), the National Institute of Mental Health (NIMH; IRP project grant number ZIA-MH002781 to A.M.W., C. Antonacci and D.S.P.; R01-MH070664 to J.P.B.; K23-MH114023 to G.A.F.; R01-MH117601 to N.J.; R00-MH117274 to A.N.K.; R01-MH70918-01A2 and R01-MH070664 to B. Milrod; R01-MH086517 and K23-MH076198 to K.L.P.; R01-MH101486 to J.W.S.; K01-MH118428 to B.S.-J.; K23-MH109983 to C.M.S.; and R01-MH116147, R01-MH121246 and R01-MH129742 to P.M.T.), the National Institutes of Health (grant number R01-MH101486 to J.A.N.), the National Institute of General Medical Sciences Center (grant number 1P20GM121312 to M.P.P.), the German Research Foundation (DFG; grant number BE 3809/8-1 to K.B.-B., KI588/14-1 and KI588/14-2 to T.K. and STR 1146/18-1 to B.S.), the Italian Ministry of Health (RF-2016-02364582 to P.B., GR-2010-2312442 and GR-2011-02348232 to C.O.), the Carlos III Health Institute (ISCIII; M.C. is funded by a 'Sara Borrell' postdoctoral contract (CD20/00189), D.P.-C. is funded by a 'PFIS' predoctoral fellowship (FI19/00251) and grant number PI18/00036 to N.C.), the Hartford HealthCare (grant number 129522 to G.J.D.), the South African Medical Research Council, Nuclear Technologies in Medicine and the Biosciences Initiative (NTeMBI) and Harry Crossley Foundation to A.G.G.D., the One Mind Baszucki-Brain Research Fund to G.A.F., the FRS-FNRS Belgian National Science Foundation (grant number 1.C.059.18F to A. Heeren), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; grant number 2013/08531-5 to A.P.J. and 2014/50917-0 to G.A.S.), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; grant number 442026/2014-5 to A.P.J. and 465550/2014-2 to G.A.S.), the Ministry of Science and ICT, South Korea (grant number NRF-2019M3C7A1032262 to S.-H.L.), the National Key R&D Program

of China (grant numbers 2022YFC2009901 and 2022YFC2009900 to S.L.), the National Natural Science Foundation of China (grant number 82120108014 to S.L.), Humboldt Foundation Friedrich Wilhelm Bessel Research Award and Chang Jiang Scholars (program number T2019069 to S.L.), the BIAL Foundation (grant number 288/20 to E.M.), the European Union's Horizon 2020 research and innovation programme (Marie Skłodowska-Curie grant number 101026595 to K.N.T.M.), the Medical Research Council New Investigator Grant (grant number MR/W005077/1 to F.M.), and the Eunice Kennedy Schriver National Institute of Child Health and Human Development (grant number K99 HD105002 to M.T.P.). A.R. is supported by a fellowship from MQ Mental Health Research and by the NIHR Oxford Health Biomedical Research Centre, the European Research Council (grant number ERC_CoG_772337 to K.R.), European Union's Seventh Framework Programme (FP7/2007-2013; grant number 337673 to G.A.S.), the McNair Foundation (MIND-DB; Veteran Health Administration grant number VHA I01CX001937 to R. Salas), the Brain and Behavior Research Foundation (NARSAD Young Investigator Grant to B.S.-J. and A.T.), and the Medical Research Council of the National Institute for Health and Care Research to B.W.

Author contributions

W.B.B. and M. Aghajani conceived and designed the study. W.B.B. and M. Aghajani collated, analyzed and interpreted the data. All authors were involved in drafting, writing and revising the paper. All authors were involved in site-level data collection and curation. All authors read and approved the final version of the paper.

Competing interests

F. Agosta is section editor of *NeuroImage: Clinical;* has received speaker honoraria from Biogen Idec, Roche and Zambon; and receives or has received research supports from the Italian Ministry of Health, AriSLA (Fondazione Italiana di Ricerca per la SLA) and the European Research Council. E.C. has received research support from the Italian Ministry of Health. T.D. has received consulting and speaker honoraria from Novartis, Teva and Hormosan unrelated to this paper. M.F. is editor-in-chief of the *Journal of Neurology* and associate editor of *Human Brain Mapping*, *Neurological Sciences* and *Radiology*; has received compensation for consulting services from Alexion, Almirall, Biogen, Merck, Novartis, Roche and Sanofi; has received compensation for speaking activities from Bayer, Biogen, Celgene, Chiesi Italia SpA, Eli Lilly, Genzyme, Janssen, Merck-Serono, Neopharmed Gentili, Novartis, Novo Nordisk, Roche, Sanofi, Takeda and Teva; has participated in advisory boards for Alexion, Biogen, Bristol-Myers Squibb, Merck, Novartis, Roche, Sanofi, Sanofi-Aventis, Sanofi-Genzyme and Takeda; has participated in scientific direction of educational events for Biogen, Merck, Roche, Celgene, Bristol-Myers Squibb, Lilly, Novartis and Sanofi-Genzyme; and receives research support from Biogen Idec, Merck-Serono, Novartis, Roche, Italian Ministry of Health, Fondazione Italiana Sclerosi Multipla and ARiSLA (Fondazione Italiana di Ricerca per la SLA). G.A.F. owns equity in Alto Neuroscience and is a consultant to Synapse Bio AI. H.J.G. has received travel grants and speaker honoraria from Fresenius Medical Care, Neuraxpharm, Servier and Janssen Cilag, and has received research funding from Fresenius Medical Care. N.J. received partial research grant support from Biogen for research unrelated to this paper. M.P.P. is an advisor to Spring Care, a behavioral health start-up, and has received royalties for an article about methamphetamine in UpToDate. P.M.T. received partial research grant support from Biogen for research unrelated to this paper. All other individually named authors in and outside of the ENIGMA-Anxiety Working Groups reported no biomedical financial interests or potential conflicts of interest.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44220-023-00173-2.

Correspondence and requests for materials should be addressed to Moji Aghajani.

Peer review information *Nature Mental Health* thanks Alice Chavanne, Ruiyang Ge and the other, anonymous reviewer(s) for their contribution to the peer review of this work.

Willem B. Bruin[1,2,3,4], Paul Zhutovsky[1,2], Guido A. van Wingen[1,2], Janna Marie Bas-Hoogendam[5,6,7], Nynke A. Groenewold[8,9], Kevin Hilbert[10,11], Anderson M. Winkler[12,13], Andre Zugman[13], Federica Agosta[14,15,16], Fredrik Åhs[17], Carmen Andreescu[18], Chase Antonacci[19,20], Takeshi Asami[21], Michal Assaf[22,23], Jacques P. Barber[24], Jochen Bauer[25], Shreya Y. Bavdekar[26], Katja Beesdo-Baum[27], Francesco Benedetti[28], Rachel Bernstein[19], Johannes Björkstrand[29], Robert J. Blair[30,31], Karina S. Blair[32], Laura Blanco-Hinojo[33,34,35], Joscha Böhnlein[36], Paolo Brambilla[37,38], Rodrigo A. Bressan[39,40], Fabian Breuer[36], Marta Cano[41,42], Elisa Canu[14], Elise M. Cardinale[43], Narcís Cardoner[41,42,44], Camilla Cividini[14], Henk Cremers[45], Udo Dannlowski[36], Gretchen J. Diefenbach[46,47], Katharina Domschke[48], Alexander G. G. Doruyter[49,50], Thomas Dresler[51,52], Angelika Erhardt[53], Massimo Filippi[14,15,16,54], Gregory A. Fonzo[55], Gabrielle F. Freitag[56], Tomas Furmark[57], Tian Ge[58,59], Andrew J. Gerber[60,61], Savannah N. Gosnell[62], Hans J. Grabe[63], Dominik Grotegerd[36], Ruben C. Gur[64,65,66], Raquel E. Gur[64,65], Alfons O. Hamm[67], Laura K. M. Han[68,69], Jennifer C. Harper[70], Anita Harrewijn[19,71], Alexandre Heeren[72], David Hofmann[73], Andrea P. Jackowski[39,40,74], Neda Jahanshad[75], Laura Jett[19], Antonia N. Kaczkurkin[76], Parmis Khosravi[13], Ellen N. Kingsley[77], Tilo Kircher[78,79], Milutin Kostic[80,81], Bart Larsen[64], Sang-Hyuk Lee[82], Elisabeth J. Leehr[36], Ellen Leibenluft[83], Christine Lochner[84], Su Lui[85], Eleonora Maggioni[38,86], Gisele G. Manfro[87,88], Kristoffer N. T. Månsson[89], Claire E. Marino[26], Frances Meeten[90], Barbara Milrod[91], Ana Munjiza Jovanovic[80,81], Benson Mwangi[92], Michael J. Myers[70], Susanne Neufang[93], Jared A. Nielsen[94,95,96,97], Patricia A. Ohrmann[98],

Cristina Ottaviani[99,100], Martin P. Paulus[101], Michael T. Perino[102], K. Luan Phan[103], Sara Poletti[28], Daniel Porta-Casteràs[41,44,104], Jesus Pujol[33,34], Andrea Reinecke[105,106], Grace V. Ringlein[19], Pavel Rjabtsenkov[26], Karin Roelofs[107], Ramiro Salas[62,108,109], Giovanni A. Salum[40,110], Theodore D. Satterthwaite[64,111], Elisabeth Schrammen[36], Lisa Sindermann[112], Jordan W. Smoller[58,59], Jair C. Soares[92], Rudolf Stark[79,113,114], Frederike Stein[78,79], Thomas Straube[73], Benjamin Straube[78,79], Jeffrey R. Strawn[115,116], Benjamin Suarez-Jimenez[26], Chad M. Sylvester[70], Ardesheer Talati[117,118], Sophia I. Thomopoulos[75], Raşit Tükel[119], Helena van Nieuwenhuizen[120], Kathryn Werwath[19], Katharina Wittfeld[63,121], Barry Wright[122,123], Mon-Ju Wu[92], Yunbo Yang[78,79], Anna Zilverstand[124], Peter Zwanzger[125,126], Jennifer U. Blackford[127,128], Suzanne N. Avery[128], Jacqueline A. Clauss[129,130], Ulrike Lueken[10,131], Paul M. Thompson[75], Daniel S. Pine[13], Dan J. Stein[132], Nic J. A. van der Wee[6,7,133], Dick J. Veltman[3] & Moji Aghajani[3,4]

[1]Department of Psychiatry, Amsterdam UMC, Location University of Amsterdam, Amsterdam, the Netherlands. [2]Amsterdam Neuroscience, Amsterdam, the Netherlands. [3]Department of Psychiatry, Amsterdam UMC, Location Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [4]Section Forensic Family and Youth Care, Institute of Education and Child Studies, Leiden University, Leiden, the Netherlands. [5]Department of Developmental and Educational Psychology, Leiden University, Leiden, the Netherlands. [6]Department of Psychiatry, Leiden University Medical Center, Leiden, the Netherlands. [7]Leiden Institute for Brain and Cognition, Leiden, the Netherlands. [8]SAMRC Unit on Child and Adolescent Health and Department of Paediatrics and Child Health, Red Cross War Memorial Children's Hospital, University of Cape Town, Cape Town, South Africa. [9]Department of Psychiatry and Mental Health, Neuroscience Institute, University of Cape Town, Cape Town, South Africa. [10]Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany. [11]Department of Psychology, Health and Medical University Erfurt, Erfurt, Germany. [12]Department of Human Genetics, University of Texas Rio Grande Valley, Brownsville, TX, USA. [13]Section on Development and Anxiety Disorders, National Institute of Mental Health Intramural Program, Bethesda, MD, USA. [14]Neuroimaging Research Unit, Division of Neuroscience, IRCCS San Raffaele Scientific Institute, Milan, Italy. [15]Neurology Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy. [16]Vita-Salute San Raffaele University, Milan, Italy. [17]Department of Psychology and Social Work, Mid Sweden University, Östersund, Sweden. [18]Department of Psychiatry, University of Pittsburgh, Pittsburgh, KS, USA. [19]Section on Development and Affective Neuroscience, National Institute of Mental Health, Bethesda, MD, USA. [20]Department of Psychology, Stanford University, Palo Alto, CA, USA. [21]Department of Psychiatry, Yokohama City University School of Medicine, Yokohama, Japan. [22]Olin Neuropsychiatry Research Center, Institute of Living, Hartford, CT, USA. [23]Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA. [24]Derner School of Psychology, Adelphi University, Garden City, NY, USA. [25]Department of Radiology, University of Münster, Münster, Germany. [26]The Del Monte Institute for Neuroscience, Department of Neuroscience, University of Rochester School of Medicine and Dentistry, Rochester, NY, USA. [27]Behavioral Epidemiology, Institute of Clinical Psychology and Psychotherapy, TUD – Dresden University of Technology, Dresden, Germany. [28]Division of Neuroscience, IRCCS San Raffaele Scientific Institute, Milan, Italy. [29]Department of Psychology, Lund University, Lund, Sweden. [30]Child and Adolescent Mental Health Centre, Copenhagen University Hospital – Mental Health Services CPH, Copenhagen, Denmark. [31]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [32]Center for Neurobehavioral Research, Boys Town National Research Hospital, Boys Town, NE, USA. [33]MRI Research Unit, Hospital del Mar, Barcelona, Spain. [34]CIBER de Salud Mental, Instituto de Salud Carlos III, Barcelona, Spain. [35]Fundació Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Barcelona, Spain. [36]Institute for Translational Psychiatry, University of Münster, Münster, Germany. [37]Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy. [38]Department of Neurosciences and Mental Health, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. [39]Department of Psychiatry, Universidade Federal de São Paulo, São Paulo, Brazil. [40]Instituto Nacional de Psiquiatria do Desenvolvimento, São Paulo, Brazil. [41]Sant Pau Mental Health Research Group, Institut d'Investigació Biomèdica Sant Pau (IIB-Sant Pau), Hospital de la Santa Creu i Sant Pau, Barcelona, Spain. [42]CIBERSAM, Carlos III Health Institute, Madrid, Spain. [43]Department of Psychology, The Catholic University of America, Washington, D. C., USA. [44]Department of Psychiatry and Forensic Medicine, School of Medicine Bellaterra, Universitat Autònoma de Barcelona, Barcelona, Spain. [45]Department of Clinical Psychology, University of Amsterdam, Amsterdam, the Netherlands. [46]Anxiety Disorders Center, The Institute of Living/Hartford Hospital, Hartford, CT, USA. [47]Yale University School of Medicine, New Haven, CT, USA. [48]Department of Psychiatry and Psychotherapy, Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany. [49]NuMeRI Node for Infection Imaging, Central Analytical Facilities, Stellenbosch University, Stellenbosch, South Africa. [50]Division of Nuclear Medicine, Faculty of Medicine and Health Sciences, Stellenbosch University, Stellenbosch, South Africa. [51]Department of Psychiatry and Psychotherapy, Tuebingen Center for Mental Health, University Hospital Tuebingen, Tuebingen, Germany. [52]LEAD Graduate School and Research Network, University of Tuebingen, Tuebingen, Germany. [53]Translational Department, Max Planck Institute for Psychiatry, Munich, Germany. [54]Neurorehabilitation Unit and Neurophysiology Service, IRCCS San Raffaele Scientific Institute, Milan, Italy. [55]Department of Psychiatry and Behavioral Sciences, The University of Texas at Austin Dell Medical School, Austin, TX, USA. [56]Center for Children and Families, Department of Psychology, Florida International University, Miami, FL, USA. [57]Department of Psychology, Uppsala University, Uppsala, Sweden. [58]Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. [59]Center for Precision Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [60]Department of Psychiatry, Columbia University Medical Center, New York, NY, USA. [61]Silver Hill Hospital, New Canaan, CT, USA. [62]Department of Psychiatry, Baylor College of Medicine, Houston, TX, USA. [63]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany. [64]Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [65]Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [66]Philadelphia Veterans Administration Medical Center, Philadelphia, PA, USA. [67]Department of Psychology, University of Greifswald, Greifswald, Germany. [68]Centre for Youth Mental Health, The University of Melbourne, Parkville, VIC, Australia. [69]Orygen, Parkville, VIC, Australia. [70]Department of Psychiatry, Washington University, St Louis, MO, USA. [71]Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, Rotterdam, the Netherlands. [72]Psychological Sciences Research Institute, Université Catholique de Louvain, Louvain-la-Neuve, Belgium. [73]Institute of Medical Psychology and Systems Neuroscience, University of Münster, Münster, Germany. [74]Department of Education, ICT and Learning, Østfold University College, Halden, Norway. [75]Imaging Genetics Center, Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. [76]Department of Psychology, Vanderbilt University, Nashville, TN, USA. [77]Child Oriented Mental Health Innovation Collaborative, Leeds and York Partnership NHS Foundation Trust, York, UK. [78]Department of Psychiatry and Psychotherapy, University of Marburg, Marburg, Germany. [79]Center of Mind, Brain and Behavior, Universities of Marburg and Giessen, Marburg/Giessen, Germany. [80]Institute of Mental Health, Belgrade, Serbia. [81]Faculty of Medicine, University of Belgrade, Belgrade, Serbia. [82]Department of Psychiatry, CHA Bundang Medical Center, CHA University, Seongnam-si, Republic of Korea. [83]Section on

Mood Dysregulation and Neuroscience, National Institute of Mental Health Intramural Program, Bethesda, MD, USA. [84]SAMRC Unit on Risk and Resilience in Mental Disorders, Department of Psychiatry, Stellenbosch University, Stellenbosch, South Africa. [85]Huaxi MR Research Center (HMRRC), Department of Radiology, West China Hospital of Sichuan University, Chengdu, China. [86]Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy. [87]Department of Psychiatry, School of Medicine, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil. [88]Anxiety Outpatient Unit, Hospital de Clinicas de Porto Alegre, Porto Alegre, Brazil. [89]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden. [90]School of Psychology, University of Sussex, Falmer, UK. [91]Albert Einstein College of Medicine, Bronx, NY, USA. [92]Louis A. Faillace, MD, Department of Psychiatry and Behavioral Sciences, The University of Texas Health Science Center at Houston, Houston, TX, USA. [93]Department of Psychiatry and Psychotherapy, Medical Faculty Heinrich-Heine-University Duesseldorf, Duesseldorf, Germany. [94]Department of Psychology, Harvard University, Cambridge, MA, USA. [95]Center for Brain Science, Harvard University, Cambridge, MA, USA. [96]Department of Psychology, Brigham Young University, Provo, UT, USA. [97]Neuroscience Center, Brigham Young University, Provo, UT, USA. [98]Department of Psychiatry, University of Münster, Münster, Germany. [99]Department of Psychology, Sapienza University of Rome, Rome, Italy. [100]Neuroimaging Laboratory, IRCCS Santa Lucia Foundation, Rome, Italy. [101]Laureate Institute for Brain Research, Tulsa, OK, USA. [102]Washington University School of Medicine, St Louis, MO, USA. [103]Department of Psychiatry and Behavioral Health, The Ohio State University, Columbus, OH, USA. [104]Mental Health Department, Unitat de Neurociència Traslacional, Parc Taulí University Hospital, Institut d'Investigació i Innovació Sanitària Parc Taulí (I3PT), Barcelona, Spain. [105]Department of Psychiatry, University of Oxford, Oxford, UK. [106]Oxford Health NHS Foundation Trust, Oxford, UK. [107]Donders Institute for Brain Cognition and Behaviour and Behavioural Science Institute, Radboud University, Nijmegen, the Netherlands. [108]Center for Translational Research on Inflammatory Diseases, Michael E. DeBakey VA Medical Center, Houston, TX, USA. [109]The Menninger Clinic, Houston, TX, USA. [110]Department of Psychiatry, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil. [111]Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA, USA. [112]Institute of Human Genetics, University of Bonn, School of Medicine and University Hospital Bonn, Bonn, Germany. [113]Department of Psychotherapy and Systems Neuroscience, Justus Liebig University Giessen, Giessen, Germany. [114]Bender Institute of Neuroimaging, Justus Liebig University Giessen, Giessen, Germany. [115]University of Cincinnati, Cincinnati, OH, USA. [116]Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. [117]Department of Psychiatry, Columbia University Irving Medical Center, New York, NY, USA. [118]Division of Translational Epidemiology, New York State Psychiatric Institute, New York, NY, USA. [119]Department of Psychiatry, Istanbul University Istanbul Faculty of Medicine, Istanbul, Turkey. [120]Department of Physics, Stony Brook University, Stony Brook, NY, USA. [121]German Center for Neurodegenerative Diseases (DZNE), Site Rostock/Greifswald, Greifswald, Germany. [122]Hull York Medical School, University of York, York, UK. [123]Child Oriented Mental Health Innovation Collaborative, Leeds and York Partnership NHS Foundation Trust, Leeds, UK. [124]Department of Psychiatry and Behavioral Sciences, University of Minnesota, Minneapolis, MN, USA. [125]Department of Psychiatry, LMU Munich, Munich, Germany. [126]kbo-Inn-Salzach-Klinikum Wasserburg, Wasserburg, Germany. [127]University of Nebraska Medical Center, Omaha, NE, USA. [128]Vanderbilt University Medical Center, Nashville, TN, USA. [129]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [130]Harvard Medical School, Boston, MA, USA. [131]German Center for Mental Health (DZPG), Site Berlin/Potsdam, Berlin, Germany. [132]SAMRC Unit on Risk and Resilience in Mental Disorders, Department of Psychiatry and Neuroscience Institute, University of Cape Town, Cape Town, South Africa. [133]Theme Neuroscience, Leiden University Medical Center, Leiden, the Netherlands. ✉e-mail: m.aghajani@fsw.leidenuniv.nl

nature portfolio

Corresponding author(s): Dr. Moji Aghajani

Last updated by author(s): 14-09-2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection, it concerns a mega-analysis of existing data (secondary data analyses).
Data analysis	Software/code for structural MRI data extraction, preprocessing, and visualization is available at http://enigma.ini.usc.edu/protocols/imaging-protocols/ https://enigma-toolbox.readthedocs.io/en/latest/index.html Regional mean cortical thickness (CT), cortical surface area (SA) and subcortical volumes (SCV), were extracted from the brain images using FreeSurfer (version 5.3/6.0). Custom analysis scripts were developed and implemented in Python (v3.9.5) with linear support vector machines (SVM) using scikit-learn (v1.0.2). These can be found at https://github.com/WillemB2104 and also are available on reasonable request to the corresponding author (M.A), although not for commercial applications.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

There are data sharing restrictions imposed by (i) ethical review boards of the participating sites, and consent documents; (ii) national and trans-national data sharing law, such as GDPR; and (iii) institutional processes, some of which require a signed MTA for limited and predefined data use. Data sharing might still be possible though, requiring submission of a detailed analysis plan to ENIGMA ANXIETY consortium. If approved, access to relevant data is provided, depending on data availability, local PI approval, and compliance with all supervening regulations.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Sex distribution is reported in Table 1, and sex-based subgroup analyses were run to assess the impact of sex on brain-based classification performance. Gender data was not collected.
Population characteristics	Population characteristics are extensively detailed in Table 1 and Table S1
Recruitment	The current study used previously collected multi-site data for secondary mega-analyses. Participants were included from three ENIGMA-Anxiety Working Groups: Panic Disorder (PD/12 sites): 112 patients/813 controls; Generalized Anxiety Disorder (GAD/16 sites): 465 patients/1,084 controls; and Social Anxiety Disorder (SAD/13 sites): 259 patients/610 controls. All participants included in the analysis were between 10-25 years of age. HC were free of past and present psychopathology and psychotropic medication use at the time of scan. Comorbid anxiety disorders (PD/GAD/SAD) were present in some patients, and their assigned primary (DSM-IV/5 anxiety disorder) diagnosis corresponds to their respective working group. Participants' demographic and clinical characteristics are summarized in Table_1. For an overview per site, see Table_S1, with more detailed sample description and data collection procedure for each participating study site provided here https://doi.org/10.1002/hbm.25100
Ethics oversight	This study concerns a mega-analysis of existing data (secondary data analyses). This study was conducted in accordance with the Declaration of Helsinki. Individual studies were previously approved by relevant local ethical review boards, with the current study falling under the guidelines of those ethical review boards for secondary use of collected data. All participants or caretakers provided written informed consent.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We aimed for the largest size possible for our mega-analyses, by pooling existing multi-site data across various sites. To our knowledge, this the largest study of its kind.
Data exclusions	Only subjects with complete clinical and structural neuroimaging data were included.
Replication	We employed fully harmonized and previously used codes for brain data extraction, preprocessing, and analyses. Some of the codes are publicly available, some are open to sharing upon reasonable request.
Randomization	No randomization was employed, as this concerns observational data previously collected/published across various sites
Blinding	Blinding was not applicable or even possible, as we are using previously collected/published data across various sites for mega-analyses

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging