1-10-2024

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

Willem B. Bruin

Paul Zhutovsky

Guido A. van Wingen

Janna Marie Bas-Hoogendam

Nynke A. Groenewold

See next page for additional authors

Recommended Citation

Bruin, W. B., Zhutovsky, P., van Wingen, G. A., Bas-Hoogendam, J. M., Groenewold, N. A., Hilbert, K., ... & Aghajani, M. (2024). Brain-based classification of youth with anxiety disorders: transdiagnostic examinations within the ENIGMA-Anxiety database using machine learning. Nature Mental Health, 2(1), 104-118. https://doi.org/10.1038/s44220-023-00173-2

Brain-Based Classification of Youth with Anxiety Disorders: Transdiagnostic and Case-Control Examinations within ENIGMA-ANXIETY using Machine Learning

Authors

Willem B. Bruin, MSc[1,2], Paul Zhutovsky, PhD[1,2], Guido A. van Wingen, PhD[1,2], Janna Marie Bas-Hoogendam, PhD[3,4,5], Nynke A. Groenewold, PhD[6,7], Kevin Hilbert, PhD[8], Anderson M. Winkler, MD, PhD[9,10], Andre Zugman, PhD[10], Federica Agosta, MD, PhD[11,12,13], Fredrik Åhs, PhD[14], Carmen Andreescu, MD[15], Chase Antonacci, MPhil[16], Takeshi Asami, MD, PhD[17], Michal Assaf, MD[18,19], Jacques P. Barber, PhD [20], Jochen Bauer, PhD[21], Shreya Y. Bavdekar, [22], Katja Beesdo-Baum, PhD[23], Francesco Benedetti, MD[24], Rachel Bernstein, BA[16], Johannes Björkstrand, PhD[25], Robert J Blair, PhD[26], Karina S. Blair, PhD[27], Laura Blanco-Hinojo, PhD[28,29,30], Joscha Böhnlein, PhD[31], Paolo Brambilla, MD, PhD[32,33], Rodrigo A. Bressan, MD, PhD[34,35], Fabian Breuer, MSc[31], Marta Cano, PhD[36,37], Elisa Canu, PhD[11], Elise M. Cardinale, PhD[38], Narcís Cardoner, MD, PhD[36,39,37], Camilla Cividini, PhD[11], Henk Cremers, PhD[40], Udo Dannlowski, MD, PhD[31], Gretchen J. Diefenbach, PhD[41,42], Katharina Domschke, MD, PhD[43], Alexander G.G. Doruyter, PhD[44,45], Thomas Dresler, PhD[46,47], Angelika Erhardt, MD[48], Massimo Filippi, MD[11,12,49,13], Gregory A. Fonzo, PhD[50], Gabrielle F. Freitag, MSc[51], Tomas Furmark, PhD[52], Tian Ge, PhD [53,54], Andrew J. Gerber, MD, PhD[55,56], Savannah N. Gosnell, PhD[57], Hans J. Grabe, MD[58], Dominik Grotegerd, PhD[31], Ruben C. Gur, PhD[59,60,61], Raquel E. Gur, MD, PhD[59,60], Alfons O. Hamm, PhD[62], Laura K.M. Han, PhD [63,64], Jennifer C. Harper, MSc[65], Anita Harrewijn, PhD[66,16], Alexandre Heeren, PhD[67], David Hofmann, PhD[68], Andrea P. Jackowski, PhD[34,35], Neda Jahanshad, PhD[69], Laura Jett, BSc[16], Antonia N. Kaczkurkin, PhD[70], Parmis Khosravi, PhD[10], Ellen N. Kingsley, MSc[71], Tilo Kircher, MD, PhD[72], Milutin Kostic, MD, PhD[73,74], Bart Larsen, PhD[59], Sang-Hyuk Lee, MD, PhD[75], Elisabeth J. Leehr, PhD[31], Ellen Leibenluft, MD[76], Christine Lochner, PhD[77], Su Lui, PhD[78], Eleonora Maggioni, PhD[79,80], Gisele G. Manfro, MD, PhD[81,82], Kristoffer N.T. Månsson, PhD[83], Claire E. Marino, BA[22], Frances Meeten, DPhil[84], Barbara Milrod, MD[85], Ana Munjiza Jovanovic, MD, PhD[73,74], Benson Mwangi, PhD[86], Michael J. Myers, BA[65], Susanne Neufang, PhD[87], Jared A. Nielsen, PhD[88,89,90,91], Patricia A Ohrmann , MD, PhD[92], Cristina Ottaviani, PhD[93,94], Martin P. Paulus, MD[95], Michael T Perino, PhD[96], K. Luan Phan, MD[97], Sara Poletti, PhD[24], Daniel Porta-Casteràs, MSc[36,98,39], Jesus Pujol, MD, PhD[28,29], Andrea Reinecke, PhD[99,100], Grace V. Ringlein, MS[16], Pavel Rjabtsenkov, BSc[22], Karin Roelofs, PhD[101], Ramiro Salas, PhD[57,102,103], Giovanni A. Salum, PhD[104,35], Theodore D. Satterthwaite, MD[59,105], Elisabeth Schrammen, MSc[31], Lisa Sindermann, Dr. rer. nat.[106], Jordan W. Smoller, MD, ScD[53,54], Jair C. Soares, MD, PhD[86], Rudolf Stark, PhD[107,108,109], Frederike Stein, PhD[72,110], Thomas Straube, PhD[68], Benjamin Straube, PhD[72], Jeffrey R. Strawn, MD[111,112], Benjamin Suarez-Jimenez, PhD[22], Chad M. Sylvester, MD, PhD[65], Ardesheer Talati, PhD[113,114], Sophia I. Thomopoulos, BA[69], Raşit Tükel, MD[115], Helena van Nieuwenhuizen, BSc[116], Kathryn Werwath, BSc[16], Katharina Wittfeld, PhD[117,58], Barry Wright, MD[118,119], Mon-Ju Wu, PhD[86], Yunbo Yang, PhD[72,110], Anna Zilverstand, PhD[120], Peter Zwanzger, MD[121,122], Jennifer U Blackford, PhD[123,124], Suzanne N Avery, PhD[125], Jacqueline A. Clauss, MD, PhD[126,127], Ulrike Lueken, PhD[8], Paul M. Thompson, PhD[69], Daniel S. Pine, MD[10], Dan J Stein, MD, PhD[128], Nic J.A. van der Wee, MD, PhD[4,5,129], Dick J. Veltman, MD, PhD[130], Moji Aghajani, PhD[131,130]

Affiliations

[1]Amsterdam UMC location University of Amsterdam, Department of Psychiatry, Meibergdreef 9, Amsterdam, The Netherlands. [2]Amsterdam Neuroscience, Amsterdam, The Netherlands. [3]Leiden University, Department of Developmental and Educational Psychology, Leiden, The Netherlands. [4]Department of Psychiatry, Leiden University Medical Center, Leiden, the Netherlands. [5]Leiden Institute for Brain and Cognition, Leiden, The Netherlands. [6]SA-MRC Unit on Child & Adolescent Health & Department of Paediatrics and Child Health, Red Cross War Memorial Children's Hospital, University of Cape Town, Cape Town, South Africa. [7]Department of Psychiatry & Mental Health, Neuroscience Institute, University of Cape Town, Cape Town, South Africa. [8]Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany. [9]Department of Human Genetics, University of Texas Rio Grande Valley, Brownsville, TX, USA. [10]Section on Development and Anxiety Disorders, National Institute of Mental Health Intramural Program, Bethesda, MD, USA. [11]Neuroimaging Research Unit, Division of Neuroscience, IRCCS San Raffaele Scientific Institute, Milan, Italy. [12]Neurology Unit, IRCCS San Raffaele Scientific Institute, Milan, Italy. [13]Vita-Salute San Raffaele University, Milan, Italy. [14]Department of Psychology and Social Work, Mid Sweden University, Östersund, Sweden. [15]University of Pittsburgh, Department of Psychiatry, Pittsburgh, USA. [16]Section on Development and Affective Neuroscience, National Institute of Mental Health, Bethesda, MD, USA. [17]Department of Psychiatry, Yokohama City University School of Medicine, Yokohama, Japan. [18]Olin Neuropsychiatry Research Center, Institute of Living, Hartford, CT, USA. [19]Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA. [20]Derner School of Psychology, Adelphi University, Garden City, NY, USA. [21]Department of Radiology, University of Münster, Münster, Germany. [22]The Del Monte Institute for Neuroscience, Department of Neuroscience, University of Rochester School of Medicine and Dentistry, Rochester, NY, USA. [23]Behavioral Epidemiology, Institute of Clinical Psychology and Psychotherapy, Technische Universität Dresden, Germany. [24]Division of Neuroscience, IRCCS San Raffaele Scientific Institute, Milan, Italy. [25]Department of Psychology, Lund University, Lund, Sweden. [26]Child and Adolescent Mental Health Centre, Mental Health Services, Capital Region of Denmark, Copenhagen, Denmark. [27]Center for Neurobehavioral Research, Boys Town National Research Hospital, Boys Town, USA. [28]MRI Research Unit, Hospital del Mar, Barcelona, Spain. [29]CIBER de Salud Mental, Instituto de Salud Carlos III, Barcelona, Spain.. [30]Fundació Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), Barcelona, Spain. [31]Institute for Translational Psychiatry, University of Münster, Münster, Germany. [32]Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy. [33]Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. [34]Department of Psychiatry, Universidade Federal de Sao Paulo, Sao Paulo, Brazil. [35]Instituto Nacional de Psiquiatria do Desenvolvimento, São Paulo, Brazil. [36]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. [37]CIBERSAM, Carlos III Health Institute, Madrid, Spain. [38]Department of Psychology, The Catholic University of America. [39]Department of Psychiatry and Forensic Medicine, School of Medicine Bellaterra, Universitat Autònoma de Barcelona, Barcelona, Spain. [40]Department of Clinical Psychology, University of Amsterdam, Amsterdam, the Netherlands. [41]Anxiety Disorders Center, The Institute of Living/Hartford Hospital, Hartford, CT, USA. [42]Yale University School of Medicine, New Haven, CT, USA. [43]Department of Psychiatry and Psychotherapy, Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany. [44]NuMeRI Node for Infection Imaging, Central Analytical Facilities, Stellenbosch University, Stellenbosch, South Africa. [45]Division of Nuclear Medicine, Faculty of Medicine and Health Sciences, Stellenbosch University, Stellenbosch, South Africa. [46]Department of Psychiatry and Psychotherapy, Tuebingen Center for Mental Health, University Hostpital Tuebingen, Germany. [47]LEAD Graduate School & Research Network, University of Tuebingen, Germany. [48]Max Planck Institute for Psychiatry, Translational Department, Munich, Germany. [49]Neurorehabilitation Unit and Neurophysiology Service, IRCCS San Raffaele Scientific Institute, Milan, Italy. [50]Department of Psychiatry and Behavioral Sciences, The University of Texas at Austin Dell Medical School, Austin, TX, USA. [51]Center for Children and Families, Department of Psychology, Florida International University, Miami, FL, USA. [52]Department of Psychology, Uppsala University, Uppsala, Sweden. [53]Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine,

Massachusetts General Hospital, Boston, MA, USA. [54]Center for Precision Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [55]Department of Psychiatry, Columbia University Medical Center, New York, NY, USA. [56]Silver Hill Hospital, New Canaan, CT, USA. [57]Department of Psychiatry, Baylor College of Medicine, Houston TX, 77030, USA. [58]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany. [59]Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [60]Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [61]Philadelphia Veterans Administration Medical Center, Philadelphia, PA 19104. [62]Department of Psychology, University of Greifswald, Germany. [63]Centre for Youth Mental Health, The University of Melbourne, Parkville, VIC, Australia. [64]Orygen, Parkville, VIC, Australia. [65]Department of Psychiatry, Washington University, St. Louis, MO, USA. [66]Department of Psychology, Education & Child Studies, Erasmus University Rotterdam, Rotterdam, The Netherlands. [67]Psychological Sciences Research Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium. [68]Institute of Medical Psychology and Systems Neuroscience, University of Münster, Germany. [69]Imaging Genetics Center, Stevens Neuroimaging & Informatics Institute, Keck School of Medicine, University of Southern California, Los Angeles, California, USA. [70]Department of Psychology, Vanderbilt University, Nashville, TN 37240, USA. [71]Child Oriented Mental Health Innovation Collaborative, Leeds and York Partnership Foundation Trust, NHS, York UK. [72]Department of Psychiatry and Psychotherapy, University of Marburg, Marburg, Germany. [73]Institute of Mental Health, Belgrade, Serbia. [74]Faculty of Medicine, University of Belgrade, Serbia. [75]Department of Psychiatry, CHA Bundang Medical Center, CHA University, 59 Yatap-ro, Bundang-gu, Seongnam-si, Gyeonggi-do, 13496, Republic of Korea. [76]Section on Mood Dysregulation and Neuroscience, National Institute of Mental Health Intramural Program, Bethesda, MD, USA. [77]SAMRC Unit on Risk & Resilience in Mental Disorders, Dept of Psychiatry, Stellenbosch University, South Africa. [78]Huaxi MR Research Center (HMRRC), Department of Radiology, West China Hospital of Sichuan University, Chengdu, China. [79]Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy. [80]Department of Neurosciences and Mental Health, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. [81]Department of Psychiatry, School of Medicine, Universidade Federal do Rio Grande do Sul, Brazil. [82]Anxiety Outpatient Unit, Hospital de Clinicas de Porto Alegre, Brazil. [83]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden. [84]School of Psychology, University of Sussex, Falmer, UK. [85]Albert Einstein College of Medicine, Bronx, NY, USA. [86]Louis A. Faillace, MD, Department of Psychiatry and Behavioral Sciences, The University of Texas Health Science Center at Houston, Houston, TX, USA. [87]Department of Psychiatry and Psychotherapy, Medical Faculty Heinrich-Heine-University Duesseldorf, Duesseldorf, Germany. [88]Department of Psychology, Harvard University, Cambridge, Massachusetts, USA. [89]Center for Brain Science, Harvard University, Cambridge, Massachusetts, USA. [90]Department of Psychology, Brigham Young University, Provo, UT, USA. [91]Neuroscience Center, Brigham Young University, Provo, UT, USA. [92]Department of Psychiatry, University of Münster, Münster, Germany. [93]Department of Psychology, Sapienza University of Rome, Italy. [94]Neuroimaging Laboratory, IRCCS Santa Lucia Foundation, Rome, Italy. [95]Laureate Institute for Brain Research, Tulsa OK, USA. [96]Washington University School of Medicine, St. Louis, MO. [97]Department of Psychiatry and Behavioral Health, The Ohio State University, Columbus, OH, USA. [98]Mental Health Department, Unitat de Neurociència Traslacional, Parc Tauli University Hospital, Institut d'Investigació i Innovació Sanitària Parc Taulí (I3PT), Barcelona, Spain. [99]Department of Psychiatry, University of Oxford. [100]Oxford Health NHS Foundation Trust. [101]Donders Institute for Brain Cognition and Behaviour and Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands. [102]Center for Translational Research on Inflammatory Diseases, Michael E DeBakey VA Medical Center, Houston Texas 77030, USA. [103]The Menninger Clinic, Houston, TX 77035, USA. [104]Department of Psychiatry, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil. [105]Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA 19104, USA. [106]Institute of Human Genetics, University of Bonn, School of Medicine and University Hospital Bonn, Bonn, Germany. [107]Department of Psychotherapy and Systems Neuroscience, Justus Liebig University Giessen, Germany. [108]Bender Institute of Neuroimaging, Justus Liebig University Giessen, Germany. [109]Center for Mind, Brain and Behavior (CMBB), University of Marburg and Justus Liebig. [110]Center for Mind, Brain and Behavior, University of Marburg, Germany. [111]University of Cincinnati, Cincinnati, Ohio, USA. [112]Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA. [113]Department of Psychiatry, Columbia University Irving Medical Center, New York, NY, USA. [114]Division of Translational Epidemiology, New York State Psychiatric Institute. [115]Department of Psychiatry, Istanbul University Istanbul Faculty of Medicine, Istanbul, Turkey. [116]Department of Physics, Stony Brook University, Stony Brook, NY, USA. [117]German Center for Neurodegenerative Diseases (DZNE), Site Rostock/ Greifswald, Greifswald, Germany. [118]Hull York Medical School, University of York, York, UK. [119]Child Oriented Mental Health Innovation Collaborative, Leeds and York Partnership Foundation Trust, NHS, Leeds UK.. [120]Department of Psychiatry and Behavioral Sciences, University of Minnesota, USA. [121]Department of Psychiatry, LMU Munich, Germany. [122]kbo-Inn-Salzach-Klinikum Wasserburg Germany. [123]University of Nebraska Medical Center, Omaha, Nebraska. [124]Vanderbilt University Medical Center, Nashville, TN. [125]Vanderbilt University Medical Center, Nashville, Tennessee. [126]Department of Psychiatry, Massachusetts General Hospital, Boston, MA. [127]Harvard Medical School. [128]SAMRC Unit on Risk & Resilience in Mental Disorders, Dept of Psychiatry & Neuroscience Institute, University of Cape Town. [129]Theme Neuroscience, Leiden University Medical Center, Leiden, the Netherlands. [130]Amsterdam UMC location Vrije Universiteit Amsterdam, Department of Psychiatry, Amsterdam, the Netherlands. [131]Leiden University, Institute of Education & Child Studies, Section Forensic Family & Youth Care, The Netherlands.

Corresponding authors: Moji Aghajani (m.aghajani@fsw.leidenuniv.nl)

Keywords: Child/Adolescent Psychiatry; Anxiety Disorders; Magnetic Resonance Imaging; Neurostructural; Machine Learning; Classification

Abstract

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 towards highly powered (i.e., big data) individual-level inferences, which are data-driven, transdiagnostic, and neurobiologically informed. Hence, we uniquely built/validated supervised neuroanatomical machine learning (ML) models for individual-level inferences, using the largest up to date neuroimaging database on youth anxiety disorders: 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 (AUC's 0.59-0.63). Classifications were driven by neuroanatomical features (cortical thickness/surface area, subcortical volumes) in fronto-striato-limbic and temporo-parietal regions. This benchmark study provides estimates on individual-level classification performances that can be realistically achieved with ML using neuroanatomical data, within a large, heterogenous, and multi-site sample of youth with anxiety disorders.

Introduction

Anxiety disorders are the most prevalent mental disorders among youth, with a life-time prevalence estimate of up to 30% (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/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), which collectively affect the perception, processing and modulation of emotionally salient information (see Figure S1). While promising, neurostructural 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 issues have prompted a shift in the field towards highly powered (i.e., big data) individual-level inferences that are data-driven, transdiagnostic, and informed by objective neurobiological measures (11,15). This shift is anticipated to provide mechanistic neurobiological insights, which would help pinpoint reliable biomarkers and potential therapeutic targets in single 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). Prior 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% (17-24). While encouraging, most of these studies 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 (i.e., age/sex distribution), clinical (i.e., medication, symptom severity, comorbidity) or technical (i.e., scanner, acquisition protocols, and diagnostic assessment) characteristics. To overcome this, large multi-site 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 hence used for the current study. The large-scale multi-site 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). While increasing sample size typically benefits ML performance and generalizability, the use of retrospectively pooled multi-site data may also complicate performance, due to the substantial heterogeneity (i.e., sample characteristics and methodology) that is inherently introduced (27-30). On the other hand, a multi-site design may give 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.

Here, we built and validated structural magnetic-resonance imaging (MRI)-based ML models based on the largest multi-site sample of neuroimaging data from young anxiety patients and healthy controls (HC) worldwide (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) (10). We focused on individuals between 10-25 years old, thus capturing the critical transition from adolescence to young adulthood, wherein these anxiety disorders first develop (3,4,31). ML classifiers were trained on brain MRI-derived cortical and subcortical gray matter features (regional cortical thickness/surface area, and subcortical volumes) to classify patients vs. HC for each anxiety disorder separately (disorder-specific), and across anxiety disorders (further referred to as transdiagnostic classification). Our hypothesis was that the classifiers would correctly classify patients vs. HC above chance-level. We aimed for entirely brain-based classifications, with additional exploratory analyses examining the influence of sex, medication usage, symptom severity and depression comorbidity subgroupings on brain-based classification performance.

Results

1. Primary Analyses

Case-control classification performances are summarized in Figure_1A, and a complete overview of results (AUC, balanced accuracy, sensitivity, specificity, permutation *p*-values) is provided in Table_S2. The best classification performance was obtained for PD vs. HC (112 PD/813 HC/12 sites) with an average AUC=0.62 (*p*=.027). None of the other disorder-specific classification performances were significantly different from chance: GAD classification (465 GAD/1,084 HC/16 sites/average AUC=0.55, *p*=.605), SAD classification (259 SAD/610 HC/13 sites/average AUC=0.57, *p*=.32), and transdiagnostic classification (823 any anxiety disorder/1,969 HC/32 sites/average AUC=0.56, *p*=.093). Classification performances obtained using LOSO-CV were similar to those obtained with repeated 5-Fold CV, in which only PD classification performed significantly above chance-level (average AUC=0.63, *p*=.003), while the SAD, GAD, and transdiagnostic classifications failed to surpass chance-level (Table_S3).

2. Exploratory Subgroup Analyses

2.1 Sex

We assessed transdiagnostic case-control classification (across ENIGMA-Anxiety subgroups) separately for males and females. Performances obtained are summarized in Figure_1B/Table_S2. Whereas above chance-level performance was achieved for male patients with any anxiety disorder vs. male HC classification (167 patients/678 HC/15 sites) (average AUC=0.63, *p*=.007), female patients with any anxiety disorder vs. female HC classification (524 patients/1,133 HC/24 sites) failed to surpass chance-level (average AUC=0.57, *p*=.218). Supplementary LOSO-CV classifications also did not surpass chance-level (Table S3).

2.2 Psychotropic Medication

Results for transdiagnostic classifications based on current psychotropic medication use are summarized in Figure_1C/Table_S2. Unmedicated patients vs. HC classification (641 patients/1,969 HC/32 sites) led to above chance-level performance (average AUC=0.59, *p*=.013), whereas medicated patients vs. HC classification (160 patients/1,307 HC/17 sites) failed to surpass chance-level (average AUC=0.59, *p*=.22). Likewise, classification of

medicated patients vs unmedicated patients (251 medicated patients/160 unmedicated patients/17 sites) failed to surpass chance-level (average AUC=0.51, *p*=.38), as was the case for supplementary LOSO-CV classifications (Table S3).

2.3 Severity

Results for transdiagnostic classifications based on severity (STAI-T: low severity≤48, high severity>48) are summarized in Figure_1D/Table_S2. Low severity patient vs. HC classification (299 patients/1,422 HC/21 sites) performance was above chance-level (average AUC=0.59, *p*=.016), while high severity patients vs. HC classification (272 patients/1,449 HC/22 sites) failed to surpass chance-level (average AUC=0.57, *p*=.305). Likewise, classification of low vs. high severity patients (299 low severity patients/270 high severity patients/21 sites) also failed to surpass chance-level (average AUC=0.52, *p*=.235), as was the case for supplementary LOSO-CV classifications (Table S3).

2.4 Comorbid Depression

Finally, we performed case-control classifications separately for patients with and without comorbid MDD. Neither of these surpassed chance-level performance with either K-Fold or LOSO-CV (see Supplementary Results, section Comorbid Depression Subgroup Analyses, for more details).

3. Feature Importance

3.1 PD Classifier

We investigated which brain regions contributed most to above chance-level classifications via sign-based consistency mapping. Four brain features emerged as most relevant for PD vs. HC classification (Figure_2/Table_S4). 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 SVM for pushing the classification *towards the PD class* given increased CT in these regions. Consistent negative signs of weights were found for CSA of the left superior frontal gyrus and SCV of the left pallidum, which would imply the opposite tendency of the SVM to assign participants to *HC class* given increased SVC/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 Supplementary Methods, section Feature Importance).

3.2 Sex, Medication, Severity Classifiers

The male patients vs. male HC transdiagnostic classification comprised seven significant features: positive signs of weights that pushed the classification *towards the patient class* were located in CT of right lateral occipital cortex and right superior frontal gyrus, and consistent negative signs of weights that pushed classification *towards HC* were found in both CT and CSA of the right superior temporal gyrus, and CSA in the *pars orbitalis*, parahippocampal gyrus and cuneus cortex in the right hemisphere (Figure_3/Table_S5).

Unmedicated patient vs. HC transdiagnostic classification comprised 21 significant features (Figure_4/Table_S6). These included consistent positive signs of weights for the right caudate SCV, CT in the right cuneus, left lingual and left superior temporal gyrus, and CSA in the left middle temporal gyrus, left entorhinal, right superior parietal and right posterior-cingulate cortex (classification towards the patient class given

increased CT/CSA). Negative signs of weights were found for the SCV of the left caudate and amygdala, CT in the left inferior parietal cortex and right middle temporal gyrus, and CSA in right inferior temporal, supramarginal and left fusiform gyrus, bilateral temporal poles, right banks superior temporal sulcus, right paracentral lobule and left medial orbital frontal and inferior parietal cortex (classification towards HC class given increased CT/CSA/SCV).

Low severity patients vs. HC transdiagnostic classification comprised six significant features (Figure_5/Table_S7). Positive signs of weights were found for CT in the right inferior parietal cortex and CSA in the right rostral anterior cingulate and left entorhinal cortex (classification towards patient class given increased CT/CSA). Consistent negative signs of weights were found for CT and CSA in the right inferior temporal gyrus, and CSA in the right *pars triangularis* (classification towards HC class given increased CT/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 multi-site ENIGMA-Anxiety Working Group, comprising a heterogeneous sample of 3,343 participants from 32 international sites. Modest classification performance was achieved for PD vs. HC (AUC=0.62), but classification performances of SAD, GAD, or any anxiety disorder (transdiagnostic) vs. HC failed to surpass chance-level significance. However, modest transdiagnostic classifications were obtained in exploratory subgroup analyses of male patients vs. male HC, unmedicated patients vs. HC, and patients with low anxiety severity vs. HC (AUC 0.59-0.63). These disorder-specific and transdiagnostic subgroup classifications were based on specific neuroanatomical features in fronto-striato-limbic and temporo-parietal regions. This study importantly shows that in a large, heterogeneous, and multi-site sample of youth with anxiety disorders, only modest classification performances can be realistically 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 multi-site 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 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 and feelings (32-34). It is embedded within the temporal lobe and the broader limbic system involved in emotional responding, and thought to play a role in PD pathophysiology (35,36). Perturbed temporal lobe GMV, specifically in the middle temporal gyrus and transverse temporal gyrus are reported in adult PD patients (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 VBM meta-analyses linking its disintegrity to PD diagnosis across age groups (37,39). Several studies also report alterations in task-based activation associated with treatment response towards 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 involved in diverse cognitive, affective and motor processes (41). Voxel-based morphometry (VBM) meta-analyses link gray matter volume

(GMV) anomalies within the lenticular nucleus to clinical anxiety in general (39), and specifically in adult PD patients (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, Severity Classifier

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 (up to 0.63 AUC, *p*'s<.05) but not for the entire sample (AUC=0.56, *p*=.093). As depicted in Figures 2-5, the multivariate neurostructural pattern behind these transdiagnostic subgroup classifications comprised of fronto-parieto-limbic regions, previously linked to anxiety and its demographic (i.e., sex) and clinical (i.e., severity/medication use) and characteristics (6,9,42-46).

Whereas transdiagnostic classification of male anxiety patients vs. male HC could be reliably achieved (AUC=0.63, *p*=.007), this was not the case for females. Such male-specific effects interestingly 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 unmedicated patients could be distinguished from HC (AUC=0.59, *p*=.013), which was not the case for classification of medicated patients vs. HC or medicated vs. unmedicated patients. One may speculate that heterogeneity in medication type, dosage, or duration among medicated patients could differentially influence brain structure and render classification more complicated. Alternatively, psychotropic medication usage may normalize some neurostructural alterations in anxiety patients (4,9), a phenomenon also seen in other psychopathologies (48,49), and hence complicate classification from HC and their unmedicated anxious peers. 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 low severity patients (as indexed by a below median STAI-trait score) could be distinguished from HC above chance-level (AUC=0.59, *p*=.016), this could not be achieved for classification of high severity patients vs. HC or high severity vs. low severity patients. While this result might be counterintuitive, it could be attributed to possible divergent brain abnormalities in high severity PD/SAD/GAD patients (e.g., larger or smaller CT/CSA/SCV depending on the disorder), or different variances in age (*p*<.001), sex (*p*=.02) and medication status (*p*<.001) we identified between low and high severity patients. The heterogeneity in the high severity patient group might be exacerbated by medication effects, as severity and medication status were found to be positively correlated (Phi=0.21, *p*<.001), which could further limit the ability of the classifier to find common brain abnormalities across these patients. Nonetheless, the significance (or lack of it, 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 tested these subgroup classifications also 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). On the other hand, 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 to one another. The lower performances obtained with LOSO-CV compared to K-Fold were also found in similar multicenter classification studies on Bipolar Disorder, Autism, Schizophrenia and Obsessive-Compulsive Disorder (28,29,50,51), and are likely due to increased heterogeneity between the training and test samples. By leaving out an entire site, 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, while 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 (obsessive-compulsive disorder (OCD), bipolar disorder (BP), major depression disorder (MDD)) that used FreeSurfer data for case-control classification in a multi-site setting, with AUCs ranging between 0.62-0.72 (27-29). Similar to our results, none of the obtained performances reached the clinical threshold of AUC>0.80 (54). These findings suggest that although there could be significant case-control differences in CT/CSA/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 (up to 0.63 AUC) translate to effect sizes (Cohen's *d*=0.47; medium effect size, see (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, such as OCD (max Cohen's *d*=−0.33 for CSA), BP (max Cohen's *d*=-0.29 for SVC), schizophrenia (SZC) (max Cohen's *d*=-0.46 for SVC), MDD (max Cohen's *d*=-0.2 for SCV, but up to −0.57 for CSA in adolescents vs matched controls), and attention deficit hyperactivity disorder (max Cohen's *d*=-0.19 for SVC) (26,55-59). However, one should be cautious to directly compare the above effect sizes, as our AUC-converted effect size was based on a multivariate method estimated through CV (out-of-sample) while the case-control effect sizes are in sample and univariate in nature. Future studies could apply more fine-grained neurostructural features (i.e., voxel-wise/vertex-wise maps), measures of brain connectivity and network function or multimodal data in combination with more sophisticated classification methods (i.e., Deep Neural Networks) (15,27). A combination of these options is postulated to improve classification performances compared to shallow ML algorithms applied to low-resolution data (15).

To our knowledge, this is the largest multi-site brain-based classification analysis of anxiety disorders among youth. Previous structural and functional MRI-based classification studies on anxiety disorders have claimed to distinguish patients from HC, with AUCs reaching over 0.80 (17-24). However, these studies relied on small samples (N<100) susceptible to unstable performance estimations during cross-validation (53,60), in which inflated performances are likely to be overrepresented 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 multi-site ENIGMA database employed 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 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 multi-site context like ENIGMA may also reveal that 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 (i.e., disorder-specific and transdiagnostic classifications, as well as exploratory subgroup 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 which site a given participant belongs to, but not when site-wise scaling was applied (Table_S8). 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. Although 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 Supplementary Discussion, section Site Effects) (66). There are other methods available to handle site-effects, such as ComBat and normative modeling, but the assumptions for these techniques (i.e., 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 & 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/overfitting were stringently confronted. The international heterogeneity and multi-site 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.

Since 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/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 (10-14 early adolescence; 15-17 middle adolescence; 18-25 late adolescence and young adults, (31,68,69). As seen in Figure_S2, there was no statistically significant association between the obtained performances and age groups (r=-0.50, p=.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/voxel-wise data or other brain-imaging modalities (i.e., function/connectivity). Incorporation of these additional data forms, and use of more sophisticated ML algorithms (i.e., deep learning) better apt in handling massively multi-modal data, might further improve classification performances reported here (15). On top of this, we would have ideally also run subgroup classifications for each anxiety disorder separately, so to gain deeper insights into disorder-specific heterogeneity, as a function of clinical/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 (See Table_S9). This precluded us from effectively probing disorder-specific heterogeneity in each anxiety disorder, something that future work should consider.

Finally, while brain data was available for every single participant, this was not the case for clinical or sociodemographic data, making it difficult to test the added value of brain-based classifiers compared to 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). While 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 to 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 (i.e. questionnaires, cognitive tests, and genetics) can lead to better diagnostic classifications (44). Moreover, the only data-driven, computational diagnostic tool currently approved by the FDA for child and adolescent psychiatry is one that actually combines neuro-bio-behavioral data (44,70). Future directions for clinical utility could also involve assessments that are specifically tailored to youth (e.g., smart phones, social media, virtual reality, gaming), along with net-benefit analyses (44). An exciting new approach within the field is one that entails multiple sequential steps informed by big data analyses with machine learning 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 towards true translational impact.

Conclusion

This study shows that in a large, heterogeneous, and multi-site sample of youth with anxiety disorders, only modest classification performances can be realistically achieved based on neuroanatomical data. That said, the obtained classification accuracies do roughly translate into a medium effect size (Cohen's d=0.47; medium effect size, see (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 multi-site consortia (27-29). This classification study sets a baseline for the ENIGMA-Anxiety Working Group, and is an important step towards 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 to more conventional diagnostic tools, such as structured clinical interviews and questionnaires.

Methods

Study Sample

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 (STROBE statement-flowchart provided in Supplement). 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. Data on ethnicity was not collected by majority of sites and could therefore not robustly be 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 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. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Image Acquisition and Processing

Structural T1-weighted 3D brain magnetic resonance imaging (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 (1.5T/3.0T); sample-specific acquisition parameters are listed elsewhere (10). 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) (72). Parcellations/segmentations were visually inspected and statistically evaluated for outliers. For each subject, SA and CT were calculated for 68 cortical Desikan-Killiany atlas-based regions (34 per hemisphere) (73). In addition, gray matter SCVs were extracted for seven structures per hemisphere (nucleus accumbens, putamen, pallidum, caudate, thalamus, amygdala, hippocampus), along with lateral ventricle volume. This yielded a total of 152 FreeSurfer features per subject for classification.

Multivariate Brain-Based Classification

1. Primary Analyses

As shown in Figure_6, fully brain-based classification tasks were performed with linear support vector machines (SVM), one of the most commonly and successfully used algorithms in the field (15,44,74). We performed disorder-specific case-control classification for each working group separately (PD vs. HC, GAD vs. HC, SAD vs. HC). We also performed a transdiagnostic classification for patients with any anxiety disorder (PD/GAD/SAD) vs. HC using pooled samples across the three working groups. 561 participants (545 HC, 5 GAD and 11 SAD patients) 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 PD).

Classification performance was evaluated using stratified 5-times-repeated-5-fold cross-validation (CV) following guidelines for the field (52,75). 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 leave-one-site-out (LOSO) CV, utilizing this supplementary validation strategy to provide complementary information to K-Fold (See Supplement for results). In each LOSO-CV fold, one site is left out for model testing and 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 area-under-the-receiver-operating-characteristic-curve (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 standard deviation of each feature on the HC from each site separately, and 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 (i.e., 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 the 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 the Supplementary Methods (section Full Dataset Case-Control Analysis). Label permutation testing with 1000 iterations (76) was used to test whether classification performance (AUC) was significantly above chance-level (α=0.05), with our hypothesis being above chance-level classification of patients vs. HC. To assess which brain regions contributed most to the classification model, we applied sign-based consistency mapping as per Gomez-Verdejo et al., 2019 (see details in Supplementary Methods, section Feature Importance) (77). We only report sign-based feature importance for classifications that passed label permutation testing.

2. 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

(State-Trait Anxiety Inventory–Trait Index: STAI-T (78); median-split on patients' scores (median=48) produced high/low severity groups) and current or lifetime comorbid major depressive disorder (MDD). Subgroup analyses were restricted to the transdiagnostic sample (any anxiety disorder vs. HC across working groups), as limited data would remain when investigating these subgroups in each disorder group separately. The classification procedure itself (i.e., site-wise scaling, imputation, undersampling and ensemble learning) was as described above. Classifications included: female patients vs. female HC, male patients vs. male HC, medicated patients vs. HC, unmedicated patients vs. HC, unmedicated patients vs. medicated patients, low severity patients (STAI-T≤48) vs. HC, high severity patients (STAI-T>48) vs. HC, low severity patients vs. high severity patients, patients with comorbid MDD vs HC, and patients without comorbid MDD vs HC.

Data availability

All data used in the current study are in principle publicly available within the ENIGMA Consortium. For this study, all site-level data was first transferred to a secure data environment (Amsterdam UMC) and subsequently subjected to the final mega-analyses, all with permission from individual sites. Some data sharing restrictions are in place though. These include restrictions imposed by (a) site-specific consent documentation, ethical review boards, and institutional processes, (b) along with national/international data-sharing legislations (e.g., GDPR). Some of these may 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. Requests or questions regarding data availability/sharing should be sent to ENIGMA-ANXIETY liaison Dr. M. Aghajani (m.aghajani@fsw.leidenuniv.nl).

Code availability

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

Acknowledgements

Study was financially supported by The Netherlands Organization for Health Research and Development-ZonMw (Research Fellowship No. 06360322210035 to M. Aghajani), Dutch Research Council NWO (SSH Open Competition No. 15810 to M. Aghajani), Leiden University Fund (Project: Youth Mental Health Meets Big Data Analytics, Grant No. LUF23075-5-306 to M. Aghajani), Leiden University Fund (Project Grant No. W213085-5 to M. Aghajani), Amsterdam Neuroscience (Amsterdam Neuroscience Alliance Grant Project to M. Aghajani & G.A. van Wingen), the Dutch Research Council NWO (Rubicon Grant No. 019.201SG.022 to J.M. Bas-Hoogendam and 452020227 to L.K.M. Han), the Carnegie Corporation of New York (to N.A. Groenewold; the statements made and views expressed are solely the responsibility of the author), the National Institute of Mental Health (NIMH; IRP Project Grant No. ZIA-MH002781 to A.M. Winkler, C. Antonacci and D.S. Pine, R01-MH070664 to J.P. Barber, K23-MH114023 to G.A. Fonzo, R01-MH117601 to N. Jahanshad, R00-MH117274 to A. Kaczkurkin, R01-MH70918-01A2 and R01-MH070664 to B. Milrod, R01-MH086517 and K23-MH076198 to K.L. Phan, R01-MH101486 to J.W. Smoller, K01-MH118428 to B. Suarez-Jimenez, K23-MH109983 to C.M. Sylvester, R01-MH116147, R01-MH121246 and R01-MH129742 to P.M. Thompson), the National Institutes of Health (NIH; Grant No. R01-MH101486 to J.A. Nielsen), the National Institute of General Medical Sciences Center (Grant No. 1P20GM121312 to M.P. Paulus), the German Research Foundation (DFG; Grant No. BE 3809/8-1 to K. Beesdo-Baum, KI588/14-1

and KI588/14-2 to T. Kircher, STR 1146/18-1 to B. Straube), the Italian Ministry of Health (RF-2016-02364582 for P. Brambila, GR-2010-2312442; GR-2011-02348232 to C. Ottaviani), the Carlos III Health Institute (ISCIII; M. Cano is funded by a 'Sara Borrell' postdoctoral contract [CD20/00189], D. Porta-Casteràs is funded by a 'PFIS' predoctoral fellowship [FI19/00251], Grant No. PI18/00036 to N. Cardoner), the Hartford HealthCare (Grant No. 129522 to G.J. Diefenbach), the South African Medical Research Council, Nuclear Technologies in Medicine and the Biosciences Initiative (NTeMBI) and Harry Crossley Foundation to A.G.G. Doruyter, the One Mind Baszucki-Brain Research Fund to G.A. Fonzo, the FRS-FNRS Belgian National Science Foundation (Grant No. 1.C.059.18F to A. Heeren), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP; Grant No. 2013/08531-5 to A.P. Jackowski, 2014/50917-0 to G.A. Salum), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; Grant No. 442026/2014-5 to A.P. Jackowski, 465550/2014-2 to G.A. Salum), the Ministry of Science and ICT, South Korea (Grant No. NRF-2019M3C7A1032262 to S-H. Lee), the National Key R&D Program of China (Grant No. 2022YFC2009901 and 2022YFC2009900 to S. Lui), the National Natural Science Foundation of China (Grant No. 82120108014 to S. Lui), Humboldt Foundation Friedrich Wilhelm Bessel Research Award and Chang Jiang Scholars (Program No. T2019069 to S. Lui), the BIAL Foundation (Grant No. 288/20 to E. Maggioni), the European Union's Horizon 2020 research and innovation programme (Marie Skłodowska-Curie Grant No. 101026595 to N.T. Månsson), the Medical Research Council New Investigator Grant (Grant No. MR/W005077/1 to F. Meeten), the Eunice Kennedy Schriver National Institute of Child Health and Human Development (Grant No. K99 HD105002 to M.T. Perino), A. Reinecke is supported by a fellowship from MQ Mental Health Research and by the NIHR Oxford Health Biomedical Research Centre, the European Research Council (Grant No. ERC_CoG_772337 to K. Roelofs, European Union's Seventh Framework Programme (FP7/2007-2013) Grant No. 337673 to G.A. Salum), the McNair Foundation (MIND-DB; Veteran Health Administration Grant No. VHA I01CX001937 to R. Salas), the Brain and Behavior Research Foundation (NARSAD Young Investigator Grant to B. Suarez-Jimenez and A. Talati), and the Medical Research Council of the National Institute for Health and Care Research to B. Wright.

Author Contributions Statement

Dr. Willem Bruin and Dr. Moji Aghajani conceived and designed the study. Dr. Willem Bruin and Dr. Moji Aghajani collated, analyzed, and interpreted the data. All authors in the author list were involved in drafting, writing, and revising the manuscript. All authors in the author list were involved in site-level data collection and curation. All authors in the author list read and approved the final version of the manuscript.

Competing Interests Statement

Prof. Dr. 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. Dr. E. Canu has received research support from the Italian Ministry of Health. Dr. T. Dresler has received consulting and speakers honoraria from Novartis, Teva and Hormosan unrelated to this manuscript. Prof. Dr. M. Filippi is Editor-in-Chief of the Journal of Neurology, Associate Editor of Human Brain Mapping, Neurological Sciences, and Radiology; received compensation for consulting services from Alexion, Almirall, Biogen, Merck, Novartis, Roche, Sanofi; 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; participation in Advisory Boards for Alexion, Biogen, Bristol-Myers Squibb, Merck, Novartis, Roche, Sanofi, Sanofi-Aventis, Sanofi-Genzyme, Takeda; scientific

direction of educational events for Biogen, Merck, Roche, Celgene, Bristol-Myers Squibb, Lilly, Novartis, Sanofi-Genzyme; he 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). Dr. G.A. Fonzo owns equity in Alto Neuroscience and is a consultant to Synapse Bio AI. Prof. Dr. H.J. Grabe has received travel grants and speakers honoraria from Fresenius Medical Care, Neuraxpharm, Servier and Janssen Cilag as well as research funding from Fresenius Medical Care. Dr. N. Jahanshad received partial research grant support from Biogen, Inc., for research unrelated to this manuscript. Prof. Dr. M.P. Paulus is an advisor to Spring Care, Inc., a behavioral health startup, he has received royalties for an article about methamphetamine in UpToDate. Prof. Dr. P.M. Thompson received partial research grant support from Biogen, Inc., for research unrelated to this manuscript. All other individually-named authors in and outside of the ENIGMA-Anxiety Working Groups reported no biomedical financial interests or potential conflicts of interest.

Figure Legends

Figure 1. Classification performances for each working group separately (1.A) and for transdiagnostic classification across working groups (1.B-1.D). Boxplots summarize scores obtained across the repeated stratified K-Fold cross-validation folds; *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. Bounds of upper whiskers represent maxima; bounds of lower whiskers represent minima; center box represents interquartile range, with mean represented by middle horizontal lines. Exact P-values, along with all classification metrics are provided in Table S2. PD = Panic Disorder, GAD = General Anxiety Disorder, SAD = Social Anxiety Disorder, HC = Healthy Control.

Sample size per classification task:

PD vs. HC: N= 112/813

GAD vs. HC: N= 465/1084

SAD vs. HC: N= 259 SAD/610

Any anxiety disorder vs. HC: N= 823/1969

Male patients vs. male HC: N= 167/678

Female patients vs. female HC: N= 524/1133

Unmedicated vs. HC: N= 641/1969

Medicated vs. HC: N= 160/1307

Unmedicated vs medicated: N= 251/160

Low severity vs. HC: N= 299/1422

High severity vs. HC: N= 272/1449

Low vs. high severity: N= 299/270

Figure 2. The -log(p) value maps characterizing significant brain regions contributing the most to panic disorder patients versus healthy controls classification as determined by sign-based consistency mapping (77)**.** Hot colors indicate consistently assigned positive weights by the SVM that drive classification towards the patients class, and cold colors indicate negative weights that drive classification towards controls. Panel A relates to subcortical volumes, panel B to cortical surface area, and panel C to cortical thickness. Z-statistic estimates from p̂-values (Importance scoring) were extracted following a T-test distribution, and used to compute two-tailed -log(p). The

figure was made with the ENIGMA Toolbox package (https://enigma-toolbox.readthedocs.io/en/latest/). Full region names and associated statistics Z, p̂, -log (p) can be found in Supplementary Table S4. LH = Left Hemisphere, RH = Right Hemisphere.

Figure 3. The -log(p) value maps characterizing significant brain regions contributing the most to male patients versus male healthy controls classification across working groups as determined by sign-based consistency mapping (77)**.** Hot colors indicate consistently assigned positive weights by the SVM that drive classification towards the patients class, and cold colors indicate negative weights that drive classification towards controls. Panel A relates to subcortical volumes, panel B to cortical surface area, and panel C to cortical thickness. Z-statistic estimates from p̂-values (Importance scoring) were extracted following a T-test distribution, and used to compute two-tailed -log(p). The figure was made with the ENIGMA Toolbox package (https://enigma-toolbox.readthedocs.io/en/latest/). Full region names and associated statistics Z, p̂, -log (p) can be found in Supplementary Table S5. LH = Left Hemisphere, RH = Right Hemisphere.

Figure 4. The -log(p) value maps characterizing significant brain regions contributing the most to unmedicated patients versus healthy controls classification across working groups as determined by sign-based consistency mapping (77)**.** Hot colors indicate consistently assigned positive weights by the SVM that drive classification towards the patients class, and cold colors indicate negative weights that drive classification towards controls. Panel A relates to subcortical volumes, panel B to cortical surface area, and panel C to cortical thickness. Z-statistic estimates from p̂-values (Importance scoring) were extracted following a T-test distribution, and used to compute two-tailed -log(p). The figure was made with the ENIGMA Toolbox package (https://enigma-toolbox.readthedocs.io/en/latest/). Full region names and associated statistics Z, p̂, -log (p) can be found in Supplementary Table S6. LH = Left Hemisphere, RH = Right Hemisphere.

Figure 5. The -log(p) value maps characterizing significant brain regions contributing the most to low severity patients versus healthy controls classification across working groups as determined by sign-based consistency mapping (77)**.** Hot colors indicate consistently assigned positive weights by the SVM that drive classification towards the patients class, and cold colors indicate negative weights that drive classification towards controls. Panel A relates to subcortical volumes, panel B to cortical surface area, and panel C to cortical thickness. Z-statistic estimates from p̂-values (Importance scoring) were extracted following a T-test distribution, and used to compute two-tailed -log(p). The figure was made with the ENIGMA Toolbox package (https://enigma-toolbox.readthedocs.io/en/latest/). Full region names and associated statistics Z, p̂, -log (p) can be found in Supplementary Table S7. LH = Left Hemisphere, RH = Right Hemisphere.

Figure 6. Simplified visual representation of the ML pipeline. 1) ML algorithm training based on labeled Freesurfer features (cortical thickness/surface area & subcortical volumes). 2) This results in a model *M* that utilizes a discriminative subset of these features (feature vector) to classify patients vs. controls (Schnack & Kahn, 2016), via an optimum separating hyperplane within the entire feature space (Schnack & Kahn, 2016). Output value y (distance to hyperplane) indicates how precise/accurate the model performs when applied to a participant's feature vector (blue=HC | red=Patients) (Schnack & Kahn, 2016). 3) False negative and false positive classifications are enclosed within overlapping parts of the separating hyperplane distribution (Schnack & Kahn, 2016). Adapted and reprinted with permission from Frontiers Media SA: Frontiers Psychiatry (Schnack & Kahn, 2016).

References

1. Zacharek SJ, Kribakaran S, Kitt ER, Gee DG (2021): Leveraging big data to map neurodevelopmental trajectories in pediatric anxiety. *Dev Cogn Neurosci*. 50:100974.
2. Merikangas KR, He JP, Burstein M, Swanson SA, Avenevoli S, Cui L, et al. (2010): 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.
3. Vigo D, Thornicroft G, Atun R (2016): Estimating the true global burden of mental illness. *Lancet Psychiatry*. 3:171-178.
4. Strawn JR, Lu L, Peris TS, Levine A, Walkup JT (2021): Research Review: Pediatric anxiety disorders - what have we learnt in the last 10 years? *J Child Psychol Psychiatry*. 62:114-139.
5. Hafstad GS, Augusti EM (2021): A lost generation? COVID-19 and adolescent mental health. *Lancet Psychiatry*. 8:640-641.
6. Strawn JR, Dominick KC, Patino LR, Doyle CD, Picard LS, Phan KL (2014): Neurobiology of Pediatric Anxiety Disorders. *Curr Behav Neurosci Rep*. 1:154-160.
7. Cosgrove VE, Rhee SH, Gelhorn HL, Boeldt D, Corley RC, Ehringer MA, et al. (2011): Structure and etiology of co-occurring internalizing and externalizing disorders in adolescents. *J Abnorm Child Psychol*. 39:109-123.
8. Costello EJ, Mustillo S, Erkanli A, Keeler G, Angold A (2003): Prevalence and development of psychiatric disorders in childhood and adolescence. *Arch Gen Psychiatry*. 60:837-844.
9. Craske MG, Stein MB, Eley TC, Milad MR, Holmes A, Rapee RM, et al. (2017): Anxiety disorders. *Nat Rev Dis Primers*. 3:17024.
10. Bas-Hoogendam JM, Groenewold NA, Aghajani M, Freitag GF, Harrewijn A, Hilbert K, et al. (2022): ENIGMA-anxiety working group: Rationale for and organization of large-scale neuroimaging studies of anxiety disorders. *Hum Brain Mapp*. 43:83-112.
11. Janssen RJ, Mourao-Miranda J, Schnack HG (2018): Making Individual Prognoses in Psychiatry Using Neuroimaging and Machine Learning. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 3:798-808.
12. Marquand AF, Kia SM, Zabihi M, Wolfers T, Buitelaar JK, Beckmann CF (2019): Conceptualizing mental disorders as deviations from normative functioning. *Mol Psychiatry*. 24:1415-1424.
13. Marquand AF, Rezek I, Buitelaar J, Beckmann CF (2016): Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies. *Biol Psychiatry*. 80:552-561.
14. Paulus MP, Thompson WK (2019): The Challenges and Opportunities of Small Effects: The New Normal in Academic Psychiatry. *JAMA Psychiatry*. 76:353-354.
15. Bzdok D, Meyer-Lindenberg A (2018): Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 3:223-230.
16. Orru G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A (2012): Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci Biobehav Rev*. 36:1140-1152.
17. Zhang W, Yang X, Lui S, Meng Y, Yao L, Xiao Y, et al. (2015): Diagnostic Prediction for Social Anxiety Disorder via Multivariate Pattern Analysis of the Regional Homogeneity. *Biomed Res Int*. 2015:763965.
18. Liu F, Guo W, Fouche JP, Wang Y, Wang W, Ding J, et al. (2015): Multivariate classification of social anxiety disorder using whole brain functional connectivity. *Brain Struct Funct*. 220:101-115.
19. Xing M, Fitzgerald JM, Klumpp H (2020): Classification of Social Anxiety Disorder With Support Vector Machine Analysis Using Neural Correlates of Social Signals of Threat. *Front Psychiatry*. 11:144.
20. Frick A, Gingnell M, Marquand AF, Howner K, Fischer H, Kristiansson M, et al. (2014): Classifying social anxiety disorder using multivoxel pattern analyses of brain function and structure. *Behav Brain Res*. 259:330-335.
21. Hilbert K, Lueken U, Muehlhan M, Beesdo-Baum K (2017): Separating generalized anxiety disorder from major depression using clinical, hormonal, and structural MRI data: A multimodal machine learning study. *Brain Behav*. 7:e00633.
22. Pantazatos SP, Talati A, Schneier FR, Hirsch J (2014): 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.
23. Lueken U, Zierhut KC, Hahn T, Straube B, Kircher T, Reif A, et al. (2016): Neurobiological markers predicting treatment response in anxiety disorders: A systematic review and implications for clinical application. *Neurosci Biobehav Rev*. 66:143-162.
24. Lueken U, Hilbert K, Wittchen HU, Reif A, Hahn T (2015): Diagnostic classification of specific phobia subtypes using structural MRI data: a machine-learning approach. *J Neural Transm (Vienna)*. 122:123-134.
25. Thompson PM, Stein JL, Medland SE, Hibar DP, Vasquez AA, Renteria ME, et al. (2014): The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav*. 8:153-182.
26. Thompson PM, Jahanshad N, Ching CRK, Salminen LE, Thomopoulos SI, Bright J, et al. (2020): 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.

27. Belov V, Erwin-Grabner T, Gonul AS, Amod AR, Ojha A, Aleman A, et al. (2022): 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. *arXiv preprint arXiv:220608122*.
28. Bruin WB, Taylor L, Thomas RM, Shock JP, Zhutovsky P, Abe Y, et al. (2020): Structural neuroimaging biomarkers for obsessive-compulsive disorder in the ENIGMA-OCD consortium: medication matters. *Transl Psychiatry*. 10:342.
29. Nunes A, Schnack HG, Ching CRK, Agartz I, Akudjedu TN, Alda M, et al. (2020): 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.
30. Schnack HG, Kahn RS (2016): Detecting Neuroimaging Biomarkers for Psychiatric Disorders: Sample Size Matters. *Front Psychiatry*. 7:50.
31. Sawyer SM, Azzopardi PS, Wickremarathne D, Patton GC (2018): The age of adolescence. *Lancet Child Adolesc Health*. 2:223-228.
32. Wang S, Zhao Y, Wang X, Yang X, Cheng B, Pan N, et al. (2021): 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.
33. Grabe HJ, Wittfeld K, Hegenscheid K, Hosten N, Lotze M, Janowitz D, et al. (2014): Alexithymia and brain gray matter volumes in a general population sample. *Hum Brain Mapp*. 35:5932-5945.
34. Ren J, Huang F, Zhou Y, Zhuang L, Xu J, Gao C, et al. (2020): 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.
35. Graeff FG, Del-Ben CM (2008): Neurobiology of panic disorder: from animal models to brain neuroimaging. *Neurosci Biobehav Rev*. 32:1326-1335.
36. Sobanski T, Wagner G, Peikert G, Gruhn U, Schluttig K, Sauer H, et al. (2010): Temporal and right frontal lobe alterations in panic disorder: a quantitative volumetric and voxel-based morphometric MRI study. *Psychol Med*. 40:1879-1886.
37. Wang X, Cheng B, Wang S, Lu F, Luo Y, Long X, et al. (2021): 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.
38. de Carvalho MR, Dias GP, Cosci F, de-Melo-Neto VL, Bevilaqua MC, Gardino PF, et al. (2010): Current findings of fMRI in panic disorder: contributions for the fear neurocircuitry and CBT effects. *Expert Rev Neurother*. 10:291-303.
39. Radua J, van den Heuvel OA, Surguladze S, Mataix-Cols D (2010): Meta-analytical comparison of voxel-based morphometry studies in obsessive-compulsive disorder vs other anxiety disorders. *Arch Gen Psychiatry*. 67:701-711.
40. Protopopescu X, Pan H, Tuescher O, Cloitre M, Goldstein M, Engelien A, et al. (2006): Increased brainstem volume in panic disorder: a voxel-based morphometric study. *Neuroreport*. 17:361-363.
41. Smith KS, Tindell AJ, Aldridge JW, Berridge KC (2009): Ventral pallidum roles in reward and motivation. *Behav Brain Res*. 196:155-167.
42. Harrewijn A, Cardinale EM, Groenewold NA, Bas-Hoogendam JM, Aghajani M, Hilbert K, et al. (2021): Cortical and subcortical brain structure in generalized anxiety disorder: findings from 28 research sites in the ENIGMA-Anxiety Working Group. *Transl Psychiatry*. 11:502.
43. Duval ER, Javanbakht A, Liberzon I (2015): Neural circuits in anxiety and stress disorders: a focused review. *Ther Clin Risk Manag*. 11:115-126.
44. Dwyer D, Koutsouleris N (2022): Annual Research Review: Translational machine learning for child and adolescent psychiatry. *J Child Psychol Psychiatry*. 63:421-443.
45. Grupe DW, Nitschke JB (2013): Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. *Nat Rev Neurosci*. 14:488-501.
46. Xie S, Zhang X, Cheng W, Yang Z (2021): Adolescent anxiety disorders and the developing brain: comparing neuroimaging findings in adolescents and adults. *Gen Psychiatr*. 34:e100411.
47. Rehbein E, Hornung J, Sundstrom Poromaa I, Derntl B (2021): Shaping of the Female Human Brain by Sex Hormones: A Review. *Neuroendocrinology*. 111:183-206.
48. Dusi N, Barlati S, Vita A, Brambilla P (2015): Brain Structural Effects of Antidepressant Treatment in Major Depression. *Curr Neuropharmacol*. 13:458-465.
49. Hajek T, Weiner MW (2016): Neuroprotective Effects of Lithium in Human Brain? Food for Thought. *Curr Alzheimer Res*. 13:862-872.
50. Abraham A, Milham MP, Di Martino A, Craddock RC, Samaras D, Thirion B, et al. (2017): Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example. *Neuroimage*. 147:736-745.
51. Nieuwenhuis M, van Haren NE, Hulshoff Pol HE, Cahn W, Kahn RS, Schnack HG (2012): Classification of schizophrenia patients and healthy controls from structural MRI scans in two large independent samples. *Neuroimage*. 61:606-612.

52. Varoquaux G, Raamana PR, Engemann DA, Hoyos-Idrobo A, Schwartz Y, Thirion B (2017): Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. *Neuroimage*. 145:166-179.
53. Flint C, Cearns M, Opel N, Redlich R, Mehler DMA, Emden D, et al. (2021): Systematic misestimation of machine learning performance in neuroimaging studies of depression. *Neuropsychopharmacology*. 46:1510-1517.
54. Botteron K, Carter C, Castellanos FX, Dickstein DP, Drevets W, Kim KL, et al. (2012): Consensus report of the APA work group on neuroimaging markers of psychiatric disorders. *Am Psychiatr Assoc*.
55. Boedhoe PSW, Schmaal L, Abe Y, Alonso P, Ameis SH, Anticevic A, et al. (2018): 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.
56. Hibar DP, Westlye LT, Doan NT, Jahanshad N, Cheung JW, Ching CRK, et al. (2018): Cortical abnormalities in bipolar disorder: an MRI analysis of 6503 individuals from the ENIGMA Bipolar Disorder Working Group. *Mol Psychiatry*. 23:932-942.
57. van Erp TG, Hibar DP, Rasmussen JM, Glahn DC, Pearlson GD, Andreassen OA, et al. (2016): Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Mol Psychiatry*. 21:547-553.
58. Schmaal L, Hibar DP, Samann PG, Hall GB, Baune BT, Jahanshad N, et al. (2017): 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.
59. Hoogman M, Bralten J, Hibar DP, Mennes M, Zwiers MP, Schweren LSJ, et al. (2017): 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.
60. Varoquaux G (2018): Cross-validation failure: Small sample sizes lead to large error bars. *Neuroimage*. 180:68-77.
61. Koppe G, Meyer-Lindenberg A, Durstewitz D (2021): Deep learning for small and big data in psychiatry. *Neuropsychopharmacology*. 46:176-190.
62. Vieira S, Liang X, Guiomar R, Mechelli A (2022): 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.
63. Eitel F, Schulz MA, Seiler M, Walter H, Ritter K (2021): Promises and pitfalls of deep neural networks in neuroimaging-based psychiatric research. *Exp Neurol*. 339:113608.
64. Woo CW, Chang LJ, Lindquist MA, Wager TD (2017): Building better biomarkers: brain models in translational neuroimaging. *Nat Neurosci*. 20:365-377.
65. Bayer JMM, Thompson PM, Ching CRK, Liu M, Chen A, Panzenhagen AC, et al. (2022): Site effects how-to and when: An overview of retrospective techniques to accommodate site effects in multi-site neuroimaging analyses. *Front Neurol*. 13:923988.
66. Winkler AM, Ridgway GR, Webster MA, Smith SM, Nichols TE (2014): Permutation inference for the general linear model. *Neuroimage*. 92:381-397.
67. Zindler T, Frieling H, Neyazi A, Bleich S, Friedel E (2020): Simulating ComBat: how batch correction can lead to the systematic introduction of false positive results in DNA methylation microarray studies. *BMC Bioinformatics*. 21:271.
68. Steinberg LD (1993): *Adolescence*. McGraw-Hill.
69. Patton GC, Olsson CA, Skirbekk V, Saffery R, Wlodek ME, Azzopardi PS, et al. (2018): Adolescence and the next generation. *Nature*. 554:458-466.
70. Abbas H, Garberson F, Liu-Mayo S, Glover E, Wall DP (2020): Multi-modular AI Approach to Streamline Autism Diagnosis in Young Children. *Sci Rep*. 10:5014.
71. Koutsouleris N, Dwyer DB, Degenhardt F, Maj C, Urquijo-Castro MF, Sanfelici R, et al. (2021): Multimodal Machine Learning Workflows for Prediction of Psychosis in Patients With Clinical High-Risk Syndromes and Recent-Onset Depression. *JAMA Psychiatry*. 78:195-209.
72. Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, et al. (2002): Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*. 33:341-355.
73. Desikan RS, Segonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. (2006): An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*. 31:968-980.
74. Claude LA, Houenou J, Duchesnay E, Favre P (2020): Will machine learning applied to neuroimaging in bipolar disorder help the clinician? A critical review and methodological suggestions. *Bipolar Disord*. 22:334-355.
75. Poldrack RA, Huckins G, Varoquaux G (2020): Establishment of Best Practices for Evidence for Prediction: A Review. *JAMA Psychiatry*. 77:534-540.
76. Ojala M, Garriga GC (2010): Permutation Tests for Studying Classifier Performance. *Journal of Machine Learning Research*. 11:1833-1863.
77. Gomez-Verdejo V, Parrado-Hernandez E, Tohka J, Alzheimer's Disease Neuroimaging I (2019): Sign-Consistency Based Variable Importance for Machine Learning in Brain Imaging. *Neuroinformatics*. 17:593-609.

78. Spielberger CD (1983): State-trait anxiety inventory for adults.