Supplemental Online Content

Chen Z, Liu X, Yang Q, et al. Evaluation of risk of bias in neuroimaging-based artificial intelligence models for psychiatric diagnosis: a systematic review. *JAMA Netw Open*. 2023;6(3):e231671. doi:10.1001/jamanetworkopen.2023.1671

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods 1: Searching strategy and RSS feed

The current study aimed to perform systematic appraisals for risk of bias (ROB) and reporting quality for neuroimaging-based AI models upon psychiatric diagnosis. We carried on literature searching at PubMed by Boolean codes, and iterated this process for searching relevant studies from 48 psychiatric categories. Valid results covered 24 psychiatric categories. Searching string codes and ICD-10-CM codes have been provided for them underneath.

Neurodevelopmental Disorders (15)
Intellectual Disabilities (15)
((((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (Intellectual Disabilities[Title/Abstract])

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/1VohRxwwapFs01yRUHPDjcJbuBXMqx60lahj5_Xq2oW0oJjgzd/?limit=15&utm_campaign=pubmed-2&fc=20220628225715

Autism Spectrum Disorder (22, F84.0)
((((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (Autism Spectrum Disorder[Title/Abstract])

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/1DYWH3zMZml5Y9Gm5i-WEb_Hia2MLqP5wejra0wh2P11gfN96i/?limit=15&utm_campaign=pubmed-2&fc=20220628230021

Attention-Deficit/Hyperactivity Disorder (25)
((((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (ADHD[Title/Abstract])

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/10__S9PoH30NANjb4xSUA__E6mPXoD_mPy4pa_W34_eoVitglx/?limit=15&utm_campaign=pubmed-2&fc=20220628230245

Specific Learning Disorder (29)
((((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR

(classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (Specific Learning Disorder[Title/Abstract]))

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/1hQS87KBiI1Vsj0-Obnu1SKYAeuT3qX69cnhOg9c-ykhHAWMP_/?limit=15&utm_campaign=pubmed-2&fc=20220628230531

Schizophrenia Spectrum and Other Psychotic Disorders (37)
Schizophrenia (42, F20.9)
((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (schizophrenia[Title/Abstract])

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/1JSuiBS--N_QZ7XDYZ2OWvBmB2FRPfFd852HC658yAB8kglIGb/?limit=15&utm_campaign=pubmed-2&fc=20220628230839

Bipolar and Related Disorders (53)
Bipolar Disorder (53)
((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (bipolar disorder[Title/Abstract])

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/1pabLar0q26DwU2MOQN9Y4P-8beBl528nlZbhQo1MNkGeNjYRG/?limit=15&utm_campaign=pubmed-2&fc=20220628231130

Depressive Disorders (77)
Major Depressive Disorder (78)
((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (major depressive disorder[Title/Abstract])

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/1jye2CWYy7Vd8H1uamxO07yg-7bnfgxqx3Id2ZR2-KMdSfPMHc/?limit=15&utm_campaign=pubmed-2&fc=20220628231433

Anxiety Disorders (97)
Social Anxiety Disorder (99, F40.10)
((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (social anxiety disorder[Title/Abstract])

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/1h5ksaKdwjS2xlJ-t-psf4gQIXT0sWMDB4mbmOC7z3QHZc9E3D/?limit=15&utm_campaign=pubmed-2&fc=20220628231914

Unspecified Anxiety Disorder (108, F41.9)
((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (unspecified anxiety disorder[Title/Abstract])

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/1jcCSTNs5uflgasbYC2ZA1anrrLMy4OdTCYynTfKWHvy5f2Y34/?limit=15&utm_campaign=pubmed-2&fc=20220628232150

General Anxiety Disorder (104, F41.1)
((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (general anxiety disorder[Title/Abstract])

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/1NosRw4HLZN0ZT0aOO2ad5ruiloRLLmG8hZQ8qXYgpFc7HBssN/?limit=15&utm_campaign=pubmed-2&fc=20220628232343

Obsessive-Compulsive and Related Disorders (109)
Obsessive-Compulsive Disorders (109, F42)
((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (Obsessive-Compulsive Disorder[Title/Abstract])

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/10SuWQ8ymABgulVsvwLdLFkl2KyCEazHKNZ

evcgBd1rB721QhP/?limit=15&utm_campaign=pubmed-2&fc=20220628232547

Trauma- and Stressor-related Disorder (119)
Posttraumatic Stress Disorder (119, F94.1)
((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (PTSD[Title/Abstract])

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/1RykBG8hFSUyFdR1rb0iMpBqRPqb28YdFQNz010LyDUJgVocBz/?limit=15&utm_campaign=pubmed-2&fc=20220628233240

Somatic Symptom and Related Disorder (135)
Somatic Symptom Disorder (135, F45.1)
((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (somatic symptom disorder[Title/Abstract])
RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/16OUJQpKnemdsQEjVKoT_8CN4t8dZJdeeEYvZrGDIh1jt7cUve/?limit=15&utm_campaign=pubmed-2&fc=20220628233448

Feeding and Eating Disorder (141)
Anorexia Nervosa (143)
((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (anorexia nervosa[Title/Abstract])

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/1VIh-1Z3_W7VXCTGj1Jqho8_xbpnh47EW0RSUX8ffWTgVwBQEp/?limit=15&utm_campaign=pubmed-2&fc=20220628233620

Bing-Eating Disorder (144, F50.2)
((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (eating disorder[Title/Abstract])

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/1Vy-VINwlaUsvWKtSqe8OGnHmeYi7j5taHeHmqOWUZ5lZmUEQB/?limit=15&utm_campaign=pubmed-2&fc=20220628234128

Sleep-Wake Disorders (151)
Insomnia Disorder (151, F51.01)
((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (insomnia disorder[Title/Abstract])

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/1bGVAOyPwMu3IIWh4rQV4ZNCC84MDibM50_Van5WfFYVxtrQ2F/?limit=15&utm_campaign=pubmed-2&fc=20220628234210

Disruptive, Impulse-Control and Conduct Disorder (183)
Conduct Disorder (183, F91.3)
((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (conduct disorder[Title/Abstract])

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/1xgjY15FSEnK08AtVIA7vJQKD-848B8UPKA_9ffLig6tkZOMTS/?limit=15&utm_campaign=pubmed-2&fc=20220629003811

Antisocial Personality Disorder (188, F60.2)
((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (antisocial personality disorder[Title/Abstract])

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/1NWuP8eFPUuryi0ODWbii-wYCw-ladba_WQmJdhkkAeC5XGYP4/?limit=15&utm_campaign=pubmed-2&fc=20220629004033

Substance-Related and Addicative Disorder (191)
Other (or Unknown) Substance Use Disorder (235)
((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT

(review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (addictive disorder[Title/Abstract])

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/1f9BUYPXGXDFlcKqoddLZBV_z4tsvMPIA5SkeJ4CSkcyjHk9CO/?limit=15&utm_campaign=pubmed-2&fc=20220629004424

Alcohol Use Disorder (195)
((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (alcohol abuse[Title/Abstract])

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/1f38Ts54ujDcS1iD53VqcHdNjk0FDb62R13N9dbXPvR8VuzQEC/?limit=15&utm_campaign=pubmed-2&fc=20220629004614

Opioid Use Disorder (216)
((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (opioid[Title/Abstract])

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/1jOCN2I_Sx6i8flFtYXnNEXnMJP9PHqfn09y-bcCERrwlq_k6W/?limit=15&utm_campaign=pubmed-2&fc=20220629004738

Personality Disorders (269)
Cluster B Personality Disorders: Borderline Personality Disorder (272, F60.3)
((((((((((((brain[Title/Abstract]) OR (neural[Title/Abstract])) OR (neuro[Title/Abstract])) AND (machine learning[Title/Abstract])) OR (multivariate pattern[Title/Abstract])) OR (classif*[Title/Abstract])) OR (biomarker[Title/Abstract])) OR (neuromarker[Title/Abstract])) OR (signature*[Title/Abstract])) OR (accurac*[Title/Abstract])) OR (sensitivity[Title/Abstract])) NOT (review[Title/Abstract])) NOT (meta analysis[Title/Abstract]))) AND (personality disorder[Title/Abstract])

RSS:https://pubmed.ncbi.nlm.nih.gov/rss/search/1zy4NRF93NhZRvOKnbY7JphTJTp929cxKFpp-643t6gWFLjgcR/?limit=15&utm_campaign=pubmed-2&fc=20220629005035

eMethods 2. Literature searching pipeline and data extraction/coding

Pipeline for searching literature has been standardized beforehand, including Boolean codes, initial review criteria and searching filters. For building broad literature pool, no eliminations codes for specific aspects in Boolean string were initially used at PubMed. Further, filters should be checked before launching a new round of literature searching, including article type, publication time, journal, species and authors. If no valid results were returned, the synonyms of searching codes (e.g., depression, depressive) would be used to try again.

Three-stage cross-validation scheme required two independent reviewers to conduct literature searching meantime in the stage 1. Tencent ™ Online Document Process (ODP) software[1] was used to record all the searching results from two independent reviewers, and labeled disparities between them for searching records automatically. Labeling results were kept blind for each other. In the stage 2, the additional reviewer was designated to tackle with these inconsistent records with previous reviewer, one-by-one, to form interim literature data pool. In the stage 3, the fourth reviewer independently performed literature searching to validate whether there were inconsistent records to interim literature data pool in the stage 2. Final literature data pool would be determined if the searching results fully replicated interim one.

Data extraction and coding mainly included two facets, with one for bibliometric metainformation and with another one for technical and reporting details for systematic appraisals.

1) Bibliometric metainformation

i. Psychiatric Category. To cover broad scope in psychiatric diagnosis, we conducted literature searching for a total of 48 categories defined by DSM[2]. Thus, the specific category for included studies was extracted and was further coded by DSM and ICD-10-CM for better clinical applicability.

ii. Fundamental information for publication. For individual analysis in the future, the title, publication time, primary affiliation for the first author, journal have been extracted. Notably, we classified affiliation into three domains, including University, Hospital and Others. The department of these affiliations were categorized into follow academic background: University: Psychology, Neuroscience, Biology, Computer Science (including medical information and biomedical engineering), Medicine, Psychiatry, and Radiology; Hospital: Pediatrics, Neuroscience, Psychiatry, Psychological Health, Computer Science (including medical information and biomedical engineering), and Radiology.

iii. Sample population. We extracted sample information and coded geospatial information for sample population of these included studies. A rare number of studies (2.01 %, 11/544) reported origins of sample population directly. In addition, a majority of studies were found to claim that "participants were recruited in local". If in this case, we labeled sample population for these studies by using geospatial locations for primary (or IRB) affiliation. Sample population for studies building AI model from open database would be recorded only once if the sample details were reported, such as ABIDE (http://fcon_1000.projects.nitrc.org/indi/abide/)[2] and ADHD-200 (http://fcon_1000.projects.nitrc.org/indi/adhd200/)[3] database.

2) Technical details

i. Model performance. We extracted quantitative metrics for model performance, such as accuracy, sensitivity, specificity and AUC.

ii. Technical parameters for AI models. We extracted data regarding algorithm, tookit, feature selection methods, sample size for each set (i.e., training, validation and testing), cross-validation (CV) scheme (e.g., k-fold CV, leave-one-subject-out CV), external validation (e.g., independent sample or cross-site validation), control analysis (e.g., data leakage), neuroimaging modality (e.g., grey matter volume, functional connectivity and EEG signals), model and codes availability, preprocessing methods (e.g., normalization, z-transformation and 0-1 standardization) and quality control.

eMethods 3. Geospatial model

We built geospatial distribution model for the globe by EasyShu ™. Spatial map for the world was defined by 251 countries or regions. Global geospatial model has been modeled by 1st fine-grained level (i.e., countries/areas). Sample population has been checked and extracted by reports in original papers. Specifically, papers may reported sample population by stating “participants were recruited in local” or “participants were enrolled in xxx University”. If in this case, sample population would be inferred into corresponding geospatial locations. Otherwise, we would email corresponding authors to enquiry details for sample population. Nevertheless,rare responses were obtained. Thus, studies with no information to infer or validate sample population were removed for building geospatial model To map sample population into geospatial model with high readability, the log-transformation was taken for sample size on each grid cell beforehand. We also provided this map with interactive availability in OSF repository (https://osf.io/m7byh/).

eMethods 4. PROBAST assessment criterion

Given that the PROBAST for AI models has been pending, we adopted PROBAST for general prediction models[4,5]. To improve applicability for assessing the ROB of psychiatry-specific features, we slightly modified the signaling question and grading criteria. Modifications along with explanations have been detailed underneath.

1) Domain 1 Participants

i. Were appropriate data sources used ?

Yes/Probably Yes: RCT, nested case-control, case-control with adjustment for baseline, cohort in open dataset, pre-registered study.

Explanation: Given the costs for neuroimaging-based techniques, it was hard to perform RCT and nest case-control design. Thus, we labeled cohort study in open dataset (e.g., ABIDE and ADHD-200) and pre-registered study as low risk of bias.

ii. Were all inclusions and exclusions of participants appropriate ?

Yes/Probably Yes: If inclusion and exclusion of participants was appropriate (,ion or poor SNR), so participants correspond to unselected participants of interest.

Explanation: Neuroimaging-based techniques were sensitive for body motions or signals detection. Thus, we highlighted these technical features for corresponding neuroimaging data as appropriate inclusion and exclusion criteria.

2) Domain 3 Outcome

i. Was the outcome determined appropriately?

Yes/Probably Yes: If a method of outcome determination has been used which is considered optimal or acceptable byDSMandICD-10.

Explanation: Ground-truth label (i.e., patient or healthy control) for psychiatric diagnosis was widely accepted to be determined by clinical psychiatric staffs referring to DSM or ICD-10.

3) Domain 4 Analysis

i. Were there a reasonable number of participants with the outcome?

Yes/Probably Yes: For model development studies, if the number of participants with the outcome relative to the number of candidate predictor parameters is $\geqslant$20 (EPV $\geqslant$20).For model validation studies, if the number of participants with the outcome is $\geqslant$200. For models basing on (functional) MRI, EEG and (functional) NIRs data, kernel-based or other techniques that can address “Curse of Dimensionality” (i.e., the number of cases is less than inputting features) was used. Explanation: Neuroimaging data generally contained ultra-high-dimension predictors (e.g., ~ 20, 000 voxles), but kernel function that used for dimension reduction could solve overfitting by mapping them into low-dimension space. Additional linear or non-linear dimension reduction techniques can be workable to address “Curse of Dimensionality” as well (e.g., PCA). In addition, the traditional EPV < 20 and n > 100 criterion that provided by PROBAST guideline are not fully suitable to judge risk of bias for these included neuroimaging-based AI models. Sample size issues have been stood for a long-lasting debate, but it increasingly converges into one line currently: requiring at least > 200 participants for

each neuroimaging-based model[6-8].

ii. Were all enrolled participants included in the analysis?
Yes/Probably Yes: If all participants enrolled in the study are included in the data analysis. Or if reasonable explanations are given for excluding participants.
Explanation: Given the technical shortcomings, some neuroimaging data may be discarded due to poor signal quality. Thus, such situation could be allowed only if the reasons is justified.

iii. Were participants with missing data handled appropriately?
No information: If a paper does not mention any information for missing data on behavioral data.
Explanation: Handling missing data is rare in neuroimaging-based studies, and it would not be accounted for "No information" in overall evaluation

iv. Were complexities in the data (e.g., censoring, competing risks, sampling of control participants) accounted for appropriately?
No/Probably No: If the case-control or nested case-control designs are used to generate sample pool with controls or explanations for how to sample from original cohort. Or if the healthy controls are selected to match case group with any controls.
Explanation: Neuroimaging studies are generally conducted by post-hoc analysis, and thus frequently used case-control and nest case-control designs. Typically speaking, they build sample pools by simplistic rules - selecting patients and matched controls randomly, which may bring out risk in data complexities.

Likewise, to fit well with neuroimaging-based studies, we modified three signaling questions for the first three domain for judge to the extent of which concerns are raised for applicability.

1) Participant Domain
i. Were included participants, the selection criteria used as well as the setting used appropriately in the primary prediction model study for clinically psychiatric diagnosis as this study claimed?
Low concern for applicability: Included participants and clinical setting were used appropriately in potential clinical practice as this study claimed.
LHighconcern for applicability: Included participants and clinical setting were used inappropriately in potential clinical practice.
Unclear concern for applicability: If relevant information about the participants and clinical setting are not reported.

2) Predictor Domain
ii. Were the definition, assessment, and timing of predictors in the primary prediction model study were used appropriately for clinically psychiatric diagnosis as it claimed?
Low concern for applicability: The definition, assessment, and timing of predictors were used appropriately in potential clinical practice as this study claimed.
High concern for applicability: The definition, assessment, and timing of predictors were used

inappropriately in potential clinical practice.
Unclear concern for applicability: If relevant information about the definition, assessment, and timing of predictors are not reported.

3) Outcome Domain

iii. Were outcome definition, timing, and method of determination in the primary prediction model study were used appropriately for clinically psychiatric diagnosis as it claimed?
Low concern for applicability: Outcome definition, timing, and method of determination were made appropriate for potential clinical practice.
High concern for applicability: Choice of outcome definition, timing, and method of outcome determination were made inappropriate for potential clinical practice.
Unclear concern for applicability: If relevant information about the outcome, timing, and method of determination is not reported.

eMethods 5. CLEAR modification

CLEAR statement was developed to guide evaluation of image-based AI model reports for dermatology by 25 items in four domains (i.e., Data, Technique, Technical assessment and Application). Thus, a portion of items fall down to outside scope of interest for us. To improve applicability for CLEAR statement, we removed three items (i.e., 8, 13, 18) and rephrased this statement by using signaling question (eTable 2). Item 8 presents "skin tone information and procedure by which skin tone was assessed", which is not scientific results for psychiatric studies. Further, as neuroimging-based technique generally contains multi-vendor detectors or multi-channel signal extractors. Thus, item 13 describing "multivendor image" has been removed. Finally, item 18 was fully specialized for dermatology prediction, and was thus removed.

To estimate the reporting quality or the risk of biased reports (ROR), we rephrased eligible 25 items into signaling questions by answers of "Report", "Partly Report", "Absence" and "Not Available". Details for how to rank the reporting quality for each item were as follow.

1) Data

i. Image types

Report: if the type of neuroimaging data are reported clearly in title/abstract or in method section;

Partly Report: if the image types are not reported directly, but can be found in method section;

Absence: no information is given for neuroimaging type

ii. Image artifacts

Report: if the quality controls (e.g., SNR, head-motion correction) are described in method section;

Partly Report: if no straightforward information for quality control, but can be found in method section (e.g., image was removed as poor quality);

Absence: no information is given;

iii. Technical acquisition details

Report: if the acquisition parameters are described in method section or in supplemental materials enough for reproducing this model;

Partly Report: if the acquisition parameters are described too rough to reproducing this model;

Absence: no information is given;

iv. Preprocessing procedure

Report: if the preprocessing procedures are described in method section or in supplemental materials enough for reproducing this model;

Partly Report: if the preprocessing are described too rough to reproducing this model;

Absence: no information is given;

v. Synthetic images made public if used

Report: if the synthetic images that used in study have been reported for name, time and reason;

Partly Report: if these public images were reported with lacking details;

Absence: no information is given;

vi. Patient-level metadata

Report: if the sociodemographic information is tabulated or is introduced in detailed;

Partly Report: if demographic information is reported quite brief;

Absence: no information is given;

vii. Potential biases that may arise from use of patient information

Report: if the clinical medical records that may influence outcome (e.g., medical history, drug-use) are reported;

Partly Report: if this study reports clinical medical records with unclear description;

Absence: no information is given;

viii. Data set partitions

Report: if the dataset partitions are reported (e.g., hold-out, cross-validation);

Partly Report: if this study reports unspecific partitions method;

Absence: no information is given;

ix. Sample sizes of training, validation and tests set

Report: if the sample size for each set is reported with certain number;

Partly Report: if the sample size is report for one set (e.g., total number);

Absence: no information is given;

x. External test set

Report: if the external validation and corresponding sample size for each set is reported with certain number;

Partly Report: if the external validation without certain sample size for each set is reported;

Absence: no information is given;

xi. Class distribution and balance

Report: if the case-control skewness (CCS) < 0.2 (CCS = $n_{patients}$ / $n_{healthy\ control}$ at $n_{patients} > n_{healthy\ control}$ or CCS = $n_{healthy\ control}$ / $n_{patients}$ at $n_{healthy\ control} > n_{patients}$) appeared;

Partly Report: if the CCS for this study > 0.2;

Absence: no information is given;

xii. Out-of-distribution images

Report: if more than one neuroimaging modality is used, it would be reported;

Partly Report: if multiple neuroimaging data is used without clear descriptions for types;

Absence: no information is given;

Overall evaluation

High reporting quality: if all the answers are “Report”;

Low reporting quality: if $\geqslant$ 1 of the answers are “Partly Report” or “Absence” unless the specific reason is given.

2) Technique

i. Labeling method

Report: if the diagnostic criteria are reported with certain contexts;

Partly Report: if diagnostic criteria are reported by claiming “followed * guideline”;

Absence: no information is given;

ii. Reference to common/accepted diagnostic label

Report: if the diagnostic criteria are reported with certain name if available (e.g., DSM, ICD-10);

Partly Report: if diagnostic criteria are reported by unclear description;
Absence: no information is given;
iii. Detailed description of algorithm development
Report: if the algorithmic details are reported in method section;
Partly Report: if algorithmic details for AI models are reported in supplemental materials so that it is hard to be followed;
Absence: no information is given;
Overall evaluation
High reporting quality: if all the answers are "Report";
Low reporting quality: if $\geqslant$ 1 of the answers are "Partly Report" or "Absence" unless the specific reason is given.

3) Technical assessment
i. How to publicly evaluate algorithm
Report: if the model availability is provided with valid webpage;
Partly Report: if codes or scripts for trained model is upon request;
Absence: no information is given;
ii. Performance measures
Report: if performance measures are reported for fundamental metrics (e.g, accuracy, sensitivity, specificity, AUC);
Partly Report: if performance measures are reported with non-quantitative form;
Absence: no information is given;
iii. Benchmarking ,technical comparison and novelty
Report: if benchmark, algorithmic comparison and novelty are discussed in discussion section;
Partly Report: if these contexts are mentioned but lacked substantial discussion;
Absence: no information is given;
iv. Bias assessment
Report: if potential bias for AI models is discussed in discussion or limitation section;
Partly Report: if the model bias are mentioned but lacked substantial discussion;
Absence: no information is given;
Overall evaluation
High reporting quality: if all the answers are "Report";
Low reporting quality: if $\geqslant$ 1 of the answers are "Partly Report" or "Absence" unless the specific reason is given.

4) Application
i. Use cases and target conditions
Report: if this study discuss the applicability for this AI model under specific conditions;
Partly Report: if this study slightly discuss applicable environment for using AI model;
Absence: no information is given;
ii. Potential impacts on the health care team and patients
Report: if this study discuss the clinical implications for this AI model;
Partly Report: if this study slightly mentions clinical potential for this AI model;

Absence: no information is given;
Overall evaluation
High reporting quality: if all the answers are “Report”;
Low reporting quality: if $\geqslant$ 1 of the answers are “Partly Report” or “Absence” unless the specific reason is given.

eMethods 6. Cohen Kappa test and heterogeneity

Cohen Kappa (*k*) test was used to assess the rating reliability across these five datapools. We conducted *k*-test for each pair of datapools on each domain (e.g., participant domain, outcome domain and analysis domain) to obtain reliability coefficients, with < 0.4 for unacceptable reliability and with 0.6-0.8 for moderate to strong reliability. It should bear in mind that no all the reliability coefficients were estimated because some datapools go against statistical prerequisites for k-tests.

Notably, reliability heterogeneity was found for both PROBAST and CLEAR assessment. Reliability coefficients for PROBAST assessment was found for a mean of 0.52 with range of 0.03 - 1.00 (median = 0.67). Further, we also found the acceptable reliability coefficients at mean of 0.79 with rang of 0.02 - 1.00 (median = 1.00). Thus, total assessment reliability was at mean of 0.68 (median = 0.96, range: 0.02 - 1.00).

For further validation on reliability towards these assessments, we publicly recruited an external assessor (LCL) who has not ever partnership with us to re-rate AI models by using PROBAST and CLEAR statements. Thirty AI models were randomly selected by external assessor for validation. Rating consistency rate (the number of consistent assessment / the total number of assessments, i.e. 30) was used to quantify the rating reliability. Results substantiated the rating reliability by showing 93.33% (28/30) consistency (median at 88%, IQR: 86 - 100%) for overall evaluation on PROBAST and 90.00% (27/30) consistency (median at 80%, IQR: 72 - 92%) on CLEAR statement.

eTable 1. Modified CLEAR signaling questions

CLEAR for Psychiatry	Answers
Domain (1) Data	
Were neuroimage types reported?	
Were reporting how to handle neuroimaging artifacts?	
Were technical acquisition details reported?	
Were preprocessing procedures reported?	
Were synthetic neuroimaging made public if used?	
Were public neuroimages referenced adequately?	
Were patient-level metadata reported enough (e.g., geographic location of patients, sex and gender distribution, ethnicity and/or race, and how it was extracted)?	
Were potential biases that arisen from use of patient information and metafata reported?	
Were dataset partitions methods reported?	
Were sample size for subsets (i.e., training, validation and test) reported?	
Were external test sets reported?	
Were class distribution and balance reported?	
Were out-of-distribution neuroimages reported?	
Domain (2) Technique	
Were labeling methods reported appropriately?	
Were references to common/accepted diagnostic labels reported?	
Were details for algorithm development reported?	
Domain (3) Technical assessment	
Were reporting how to publicly evaluate algorithm?	
Were model performance measures reported?	
Were benchmarking, technical comparison and novelty discussed?	
Were bias assessment discussed?	
Domain (4) Application	
Were use cases and target conditions discussed?	
Were potential impacts on the health care team and patients introduced?	
Overall evaluation	

eTable 2. Academic background distribution for included studies, we inferred the affiliation for first author into academic training background, e.g., Authors with affiliation of department of psychology would be assumed for scientific training in psychological domain.

Academic background	No. of models (n = 555), No. (%)
Computer and data science	268 (48.2)
Psychiatry	142 (25.6)
Neuroscience	96 (17.3)
Psychology	17 (3.1)
Radiology	13 (2.3)
Others	12 (2.2)
Biology	4 (0.7)
Medicine	2 (0.4)
Pediatrics	1 (0.2)

eTable 3. Distribution for psychiatric categories in existing AI models. Psychiatric categories were determined by DSM criteria.

Psychiatric Category	No. of models (N = 555), No. (%)
Schizophrenia	141 (25.4)
Autism Spectrum Disorder	119 (21.4)
Attention-Deficit/Hyperactivity Disorder	94 (16.9)
Major Depressive Disorder	76 (13.7)
Bipolar Disorder	35 (6.3)
Alcohol Use Disorder	17 (3.1)
Obsessive-Compulsive Disorders	17 (3.1)
Specific Learning Disorder	11 (2.0)
Posttraumatic Stress Disorder	10 (1.8)
Social Anxiety Disorder	5 (0.9)
Intellectual Disability	4 (0.7)
Other (or Unknown) Substance Use Disorder	4 (0.7)
Anorexia Nervosa	3 (0.5)
Antisocial Personality Disorder	3 (0.5)
Borderline Personality Disorder	3 (0.5)
Conduct Disorder	3 (0.5)
Unspecified Anxiety Disorder	3 (0.5)
Insomnia Disorder	2 (0.4)
Bing-Eating Disorder	1 (0.2)
Disruptive Behavior Disorders	1 (0.2)
General Anxiety Disorder	1 (0.2)
Opioid Use Disorder	1 (0.2)
Somatic Symptom Disorder	1 (0.2)

eTable 4. Distribution for AI algorithms in these models. A portion of studies possessed the absence of reporting what algorithms (classifiers), and thus were removed.

Algorithm (Classifiers)	No. of models (N = 555), No. (%)
SVM	292 (52.6)
CNN	49 (8.8)
RF(Random Forest)	20 (3.6)
GPC	16 (2.9)
ANN	14 (2.5)
LDA	14 (2.5)
Clustering	10 (1.8)
DNN	10 (1.8)
LR	13 (2.3)
GNN	6 (1.1)
XGBoost	6 (1.1)
KNN	5 (0.9)
LASSO	5 (0.9)
self-made	5 (0.9)
Decision tree	4 (0.7)
probabilistic neural network (PNN)	4 (0.7)
RVM	4 (0.7)
DBN	3 (0.5)
RNN	3 (0.5)
LSTM	3 (0.5)
Ensemble	3 (0.5)
DTL	2 (0.4)
DL	2 (0.4)
Elastic Net	2 (0.4)
CDAE	2 (0.4)
GAN	2 (0.4)
GBDT	2 (0.4)
MKL	2 (0.4)
MLP	2 (0.4)
QDA	2 (0.4)
SLR	2 (0.4)
3D CSResNet-18	1 (0.2)
AlexNet	1 (0.2)
ANFIS	1 (0.2)
ASD-DiagNet	1 (0.2)
classifier alogorithm	1 (0.2)
CNN+LSTM	1 (0.2)
DANS	1 (0.2)

Algorithm (Classifiers)	No. of models (N = 555), No. (%)
Deep autoencoder	1 (0.2)
dGCN	1 (0.2)
DL-EEGNet	1 (0.2)
DRBM	1 (0.2)
DSL	1 (0.2)
EBT	1 (0.2)
EIIC	1 (0.2)
ELM	1 (0.2)
EML	1 (0.2)
EMPaSchiz	1 (0.2)
GMM	1 (0.2)
HC	1 (0.2)
ISOMAP	1 (0.2)
KDA	1 (0.2)
L1-SCCA	1 (0.2)
LMMP	1 (0.2)
Locally Linear Embedding	1 (0.2)
Modified Adaboost classification	1 (0.2)
Multistage algorithm	1 (0.2)
Multistage classification algorithm	1 (0.2)
Multivariate machine learning method	1 (0.2)
NB	1 (0.2)
Neural network	1 (0.2)
PBL-McRBFN	1 (0.2)
Penalized Regression Model	1 (0.2)
Radial Basis Function Neural Network (RBFNN)	1 (0.2)
RBF-neural network	1 (0.2)
ResNet-50	1 (0.2)
RIM	1 (0.2)
STM	1 (0.2)
Symmetrical Uncertainty (SU)	1 (0.2)
TE-HI-GCN	1 (0.2)
Transductive learning	1 (0.2)
UMAX	1 (0.2)
ZIPMRM	1 (0.2)

eTable 5. Distribution for toolkit of these AI models.

Toolkit	No. of models (N = 555), No. (%)
Self-made	147 (26.5)
Not reported	124 (22.3)
LIBSVM	111 (20.0)
MATLAB	61 (11.0)
R packages	23 (4.1)
Pronto	21 (3.8)
Scikit-learn	16 (2.9)
Keras	10 (1.8)
PROBID	6 (1.1)
Python	6 (1.1)
Weka	5 (0.9)
NeuroMiner	4 (0.7)
SPSS	3 (0.5)
Tensorflow	3 (0.5)
Neuro-Learn	3 (0.4)
CosMo MVPA	1 (0.2)
DTREG	1 (0.2)
In-house MVPA toolbox	1 (0.2)
LG‐MMC	1 (0.2)
LSSVM Matlab toolbox	1 (0.2)
MALINI	1 (0.2)
Mania	1 (0.2)
MATLAB GUI	1 (0.2)
MVPANI	1 (0.2)
none	1 (0.2)
SAS	1 (0.2)
SLEP	1 (0.2)
Xbrain	1 (0.2)

eTable 6. Distribution for future selection methods of these AI models. Likewise, some papers lacked reports for how to do feature selection, and were thus discarded for statistics.

Feature selection	No. of models (N = 555), No. (%)
No feature selection	154 (27.7)
PCA	41 (7.4)
RFE	34 (6.1)
LASSO	24 (4.3)
ICA	19 (3.4)
Kendall tau rank	8 (1.4)
MRMR	8 (1.4)
Fisher z score	7 (1.3)
ROI	7 (1.3)
F-score	6 (1.1)
AE	5 (0.9)
L2	5 (0.9)
SVM	5 (0.9)
Discrete wavelet transforms	4 (0.7)
mRMR	4 (0.7)
Rank-based feature selection	4 (0.7)
Aprior	3 (0.5)
CV	3 (0.5)
Dual Regression ICA	3 (0.5)
Elastic Net	3 (0.5)
Forward stepwise analyses	3 (0.5)
GLM	3 (0.5)
Graph-Based Feature Selection	3 (0.5)
Max-pooling	3 (0.5)
ReliefF	3 (0.5)
sequential forward feature selection (SFFS)	3 (0.5)
SVD	3 (0.5)
Wrapping	3 (0.5)
Boruta	2 (0.4)
Ensemble feature selection	2 (0.4)
Genetic algorithm	2 (0.4)
K-S test	2 (0.4)
LDA	2 (0.4)
MVA	2 (0.4)
N2EN	2 (0.4)
ResNet50	2 (0.4)
SB	2 (0.4)
2-stage PCA	1 (0.2)
2D convolution kernel	1 (0.2)

Feature selection	No. of models (N = 555), No. (%)
AM_FM	1 (0.2)
ApEn	1 (0.2)
AR	1 (0.2)
AVLSAC	1 (0.2)
Backward elimination	1 (0.2)
Block	1 (0.2)
BSL	1 (0.2)
BWAS	1 (0.2)
CART	1 (0.2)
CBAN	1 (0.2)
CCA+Pearson graph matching sparse group lasso	1 (0.2)
CFS reduction	1 (0.2)
Clustering	1 (0.2)
COMPARE	1 (0.2)
COI	1 (0.2)
Consensuse	1 (0.2)
Cosine algorithm	1 (0.2)
CRF-based dimension reduction algorithm	1 (0.2)
DFA	1 (0.2)
DFS	1 (0.2)
Dimension Reduction	1 (0.2)
Distance	1 (0.2)
DoG	1 (0.2)
Eeset	1 (0.2)
Elman Neural Network	1 (0.2)
EMB	1 (0.2)
Empirical mode decomposition	1 (0.2)
Extra-Trees	1 (0.2)
F-TQWT	1 (0.2)
Gactor-based feature extraction	1 (0.2)
FAE	1 (0.2)
FBM	1 (0.2)
FDR	1 (0.2)
Filter Bank Common Spatial Patterns	1 (0.2)
FSV	1 (0.2)
GABM	1 (0.2)
GAD	1 (0.2)
GGLM	1 (0.2)
Greedy	1 (0.2)
Grey level co-occurrence matrix	1 (0.2)
HDPAB	1 (0.2)

Feature selection	No. of models (N = 555), No. (%)
HOG	1 (0.2)
ICA-ICN	1 (0.2)
Information Gain	1 (0.2)
Inter-ICN	1 (0.2)
Kernel PCA	1 (0.2)
Kernel Principal Component Analysis	1 (0.2)
KW test	1 (0.2)
LAAM	1 (0.2)
Latent variable	1 (0.2)
LBP-TOP	1 (0.2)
LICA	1 (0.2)
LLE	1 (0.2)
Low Frequency Components	1 (0.2)
MC-NFE	1 (0.2)
MDA	1 (0.2)
MDS	1 (0.2)
MLP	1 (0.2)
MobileNet2	1 (0.2)
Modified multiscale entropy	1 (0.2)
Motif configurations	1 (0.2)
mSVM-RFE	1 (0.2)
Multiple-task logistic regression	1 (0.2)
MVAR model	1 (0.2)
Network sparsity	1 (0.2)
NSD	1 (0.2)
JDA	1 (0.2)
PDA	1 (0.2)
PDF-FS	1 (0.2)
Peak	1 (0.2)
Pegosos	1 (0.2)
pLDA	1 (0.2)
PLI	1 (0.2)
pre-trained	1 (0.2)
PS assessment	1 (0.2)
PSCR	1 (0.2)
R-SFM	1 (0.2)
RADACC	1 (0.2)
Random forest–based feature selection	1 (0.2)
RBF	1 (0.2)
RNN	1 (0.2)
ROC	1 (0.2)
RVDM	1 (0.2)

Feature selection	No. of models (N = 555), No. (%)
SBE	1 (0.2)
Semi-supervised	1 (0.2)
Sequential minimal optimization (SMO)	1 (0.2)
SFS	1 (0.2)
Short time series selection	1 (0.2)
SNV+PCA	1 (0.2)
Sparse logistic regression (SLR)	1 (0.2)
Stepwise analysis	1 (0.2)
Stracking	1 (0.2)
Symmetrical Uncertainty (SU)	1 (0.2)
Task compasion	1 (0.2)
The multi scale ranked organizing maps (MS-ROM)	1 (0.2)
Top-k-feature-set	1 (0.2)
Units representing features	1 (0.2)
VGG16 model	1 (0.2)
Weights	1 (0.2)

eTable 7. Distribution for cross-validation regimes of these AI models.

Cross-validation regimes	No. of models (N = 555), No. (%)
LOSOCV	199 (35.9)
10-fold CV	167 (30.1)
5-fold CV	43 (7.8)
Not report	32 (5.8)
Hold-out	28 (5.1)
k-fold CV	24 (4.3)
LOsiteOCV	17 (3.1)
LOPOCV	12 (2.2)
Nested 10-fold CV	11 (2.0)
Nested LOSOCV	10 (1.8)
Nested k-fold CV	6 (1.1)
Other	2 (0.4)
leave-one-run-out	1 (0.2)
Nested 5-fold CV	1 (0.2)
Precross-validation-training	1 (0.2)
Stratified random sampling	1 (0.2)

eTable 8. Distribution for neuroimaging-based data modality of these AI models.

Data modality	No. of models (N = 555), No. (%)
RSFC	190 (34.2)
EEG	130 (23.4)
Fusion	107 (19.3)
Task-related fMRI	33 (6.0)
DTI	29 (5.2)
GMV	26 (4.7)
fNIRS	12 (2.2)
MEG	9 (1.6)
fALFF	3 (0.5)
ReHo	3 (0.5)
GMV+DTI	2 (0.4)
OCT	2 (0.4)
ADC	1 (0.2)
ASL-CBF	1 (0.2)
ASL-MRI	1 (0.2)
Brain Entropy	1 (0.2)
CSF	1 (0.2)
CT	1 (0.2)
fMRI-PSD	1 (0.2)
fMRI-PS	1 (0.2)
SPECT	1 (0.2)

eTable 9. Distribution for preprocessing methods of these AI models.

Preprocessing	No. of models (N = 555), No. (%)
None	314 (56.6)
Normalization	161 (29.0)
Fisher Z	54 (9.7)
Scaling	9 (1.6)
0-1 Normalization	3 (0.5)
Standardization	3 (0.5)
L1 or L2	2 (0.4)
Regressed Residual	2 (0.4)
LeFMSF	1 (0.2)
Bass-filter	1 (0.2)
Centering	1 (0.2)
GSP	1 (0.2)
HOG	1 (0.2)
Median	1 (0.2)
PLI	1 (0.2)

eReferences

1. Tencent Corp. Homepage. https://docs.qq.com/
2. Di Martino A, Yan CG, Li Q, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol Psychiatry. 2014;19(6):659-67.
3. The ADHD-200 Consortium: A Model to Advance the Translational Potential of Neuroimaging in Clinical Neuroscience. Front Syst Neurosci. 2012;6:62.
4. Collins GS, Dhiman P, Andaur Navarro CL, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ open. 2021;11(7):e048008.
5. Daneshjou R, Barata C, Betz-Stablein B, et al. Checklist for Evaluation of Image-Based Artificial Intelligence Reports in Dermatology: CLEAR Derm Consensus Guidelines From the International Skin Imaging Collaboration Artificial Intelligence Working Group. JAMA dermatology. 2022;158(1):90-96.
6. Flint C, Cearns M, Opel N, et al. Systematic misestimation of machine learning performance in neuroimaging studies of depression. Neuropsychopharmacology. Jul 2021;46(8):1510-1517. doi:10.1038/s41386-021-01020-7
7. Kambeitz, J., Cabral, C., Sacchet, M. D., Gotlib, I. H., Zahn, R., Serpa, M. H., Walter, M., Falkai, P., & Koutsouleris, N. (2018). Reply to: Sample Size, Model Robustness, and Classification Accuracy in Diagnostic Multivariate Neuroimaging Analyses. Biological psychiatry, 84(11), e83 – e84. https://doi.org/10.1016/j.biopsych.2018.01.023
8. Varoquaux G, Raamana PR, Engemann DA, Hoyos-Idrobo A, Schwartz Y, Thirion B. Assessing and tuning brain decoders: Cross-validation, caveats, and guidelines. Neuroimage. 2017/01/15/ 2017;145:166-179. doi:https://doi.org/10.1016/j.neuroimage.2016.10.038