RESEARCH ARTICLE

Open Access

Sampling inequalities affect generalization of neuroimaging-based diagnostic classifiers in psychiatry

Zhiyi Chen[1,2*], Bowen Hu[2], Xuerong Liu[1], Benjamin Becker[3,4], Simon B. Eickhoff[5], Kuan Miao[1], Xingmei Gu[1], Yancheng Tang[6], Xin Dai[2], Chao Li[7], Artemiy Leonov[8], Zhibing Xiao[9], Zhengzhi Feng[1], Ji Chen[10,11*] and Hu Chuan-Peng[12]

Abstract

Background The development of machine learning models for aiding in the diagnosis of mental disorder is recognized as a significant breakthrough in the field of psychiatry. However, clinical practice of such models remains a challenge, with poor generalizability being a major limitation.

Methods Here, we conducted a pre-registered meta-research assessment on neuroimaging-based models in the psychiatric literature, quantitatively examining global and regional sampling issues over recent decades, from a view that has been relatively underexplored. A total of 476 studies ($n = 118,137$) were included in the current assessment. Based on these findings, we built a comprehensive 5-star rating system to quantitatively evaluate the quality of existing machine learning models for psychiatric diagnoses.

Results A global sampling inequality in these models was revealed quantitatively (sampling Gini coefficient (G) = 0.81, $p < .01$), varying across different countries (regions) (e.g., China, $G = 0.47$; the USA, $G = 0.58$; Germany, $G = 0.78$; the UK, $G = 0.87$). Furthermore, the severity of this sampling inequality was significantly predicted by national economic levels ($\beta = -2.75$, $p < .001$, $R^2_{adj} = 0.40$; $r = -.84$, 95% CI: $-.41$ to $-.97$), and was plausibly predictable for model performance, with higher sampling inequality for reporting higher classification accuracy. Further analyses showed that lack of independent testing (84.24% of models, 95% CI: 81.0–87.5%), improper cross-validation (51.68% of models, 95% CI: 47.2–56.2%), and poor technical transparency (87.8% of models, 95% CI: 84.9–90.8%)/availability (80.88% of models, 95% CI: 77.3–84.4%) are prevailing in current diagnostic classifiers despite improvements over time. Relating to these observations, model performances were found decreased in studies with independent cross-country sampling validations (all $p < .001$, $BF_{10} > 15$). In light of this, we proposed a purpose-built quantitative assessment checklist, which demonstrated that the overall ratings of these models increased by publication year but were negatively associated with model performance.

Conclusions Together, improving sampling economic equality and hence the quality of machine learning models may be a crucial facet to plausibly translating neuroimaging-based diagnostic classifiers into clinical practice.

Keywords Psychiatric machine learning, Diagnostic classification, Meta-analysis, Neuroimaging, Sampling inequalities

*Correspondence:
Zhiyi Chen
chenzhiyi@tmmu.edu.cn
Ji Chen
ji.chen@zju.edu.cn
Full list of author information is available at the end of the article

Chen *et al. BMC Medicine*   (2023) 21:241

Page 2 of 29

Background

Machine learning (ML) models have been extensively utilized for classifying patients with mental illness to aid in clinical decision-making [1, 2]. By building machine learning models that are trained from neuroimaging-based features, the diagnostic decision could be more accurate and reliable with the aid of these objective and high-dimensional biomarkers [3, 4]. Furthermore, given the multivariate nature of brain features, machine learning techniques could capture the whole neural pattern across high-volume dependent voxels for revealing pathophysiological signatures of these disorders, while individualized prediction of machine learning models in the neuroimaging-based ML models also facilitates to address the increasing needs of precision psychiatry [5, 6]. Despite considerable efforts devoted to this end, the translation of machine learning classification for diagnostic and treatment recommendation into clinical practice remains challenging [7]. This is partly due to the poor generalizability of particular these neuroimaging-based classifiers, which are often optimized within a specific sample to incur failure of generalizing to diagnose unseen patients in new samples [8–10]. Although these classifiers can be trained to achieve a desirably high accuracy in a specific cohort, they are not representative of a more general population across medical centers, geographic regions, socioeconomic statuses, and ethnic groups [11, 12]. Moreover, persisting concerns over generalizability imply potential sampling biases despite the substantially increased size of data over recent decades [13].

As promising noninvasive, in vivo techniques (e.g., magnetic resonance imaging, MRI; electroencephalogram, EEG; positron emission computed tomography, PET), they provide unique opportunities to assess brain structure, function, and metabolic anomalies for revealing the pathophysiological signatures of these psychiatric disorders as intermediate phenotype, and hence fueled the enthusiasm in these machine learning diagnostic models [9, 14]. In addition, with the huge developments of big-data sharing initiatives (e.g., UK Biobank, Alzheimer's Disease Neuroimaging Initiative), the diagnostic studies utilizing neuroimaging-based methods for classifying psychiatric conditions have seen a remarkable proliferation at an unprecedented speed over the recent decades [9, 15]. Despite these technical merits and promising research insights, these approaches are nonetheless cost, somewhat non-scalable, and are mostly not readily available or accessible in low-income countries and regions, especially the high-field MRI and PET for neural system mapping. In this vein, probing into why and how the sampling bias and relevant factors impeded the generalizability could be a potent avenue prompting translations of these neuroimaging-based machine learning models into clinical actions. However, comprehensive knowledge about the degree of such sampling issues and what relevant factors incur poor generalizability in these models is still scarce.

The importance of replication in generalizing scientific conclusions has been increasingly stressed, and a "replication crisis" has been discussed for several decades within or beyond psychological science: multiple experimental findings fail to be replicated and generalized across populations and contexts [16, 17]. One possible underlying reason may be that the available data was primarily and predominantly drawn from WEIRD (western, educated, industrialized, rich, and democratic) societies, which mirrors a typical sampling bias [18, 19]. Specifically, in 2008, 96% studies on human behavior relied on samples from WEIRD counties, with the remaining 82% of global population being largely ignored [20, 21]. Recently, we have conducted a systematic appraisal for neuroimaging-based machine learning models in the psychiatric diagnosis by using PTOBAST (Prediction model Risk Of Bias ASsessment Tool) criterion. Results demonstrated that 83.1% of these models are at high risk of bias (ROB), and further indicate a biased distribution of sampled populations [22]. Despite these descriptive evidences, there have been no quantitative analyses conducted to clearly illustrate the extent of sampling biases at a global or regional level [22]. And what's more, the long-lasting discussion regarding the association between the regional economic level and these sampling biases remains uncertain, and requires reliable statistical evidence for clarification [22, 23]. Examining the status quo of sampling biases is particularly important for psychiatric neuroimaging-based classifiers as generalizability is critical for translating models into clinical actions [23, 24]. Patient groups, compared with non-clinical or healthy entities, are far more heterogeneous due to high inter-individual variability in psychopathology [25, 26]. This is affected not only by genetics, but environment, a broad sense covering socioeconomic status, family susceptibility, and living environment [27, 28]. Therefore, developing a generally applicable model remains challenging, as the issues raised by sampling biases may further compound poor generalizability in psychiatric classification experiments.

Apart from the generalization failures due to sampling bias, there are other pitfalls to cause overfitting as the results of heedless or intended analysis optimization. Overfitting accompanied by accuracy inflation in machine learning models refers that the results are only valid within the data used for optimization but can hardly generalize to other data drawn from the same distribution [29, 30]. In support of this notion, a recent large-scale methodological overview indicated that 87% of

machine learning models for clinical prediction exhibited a high risk of bias (ROB) for overfitting, particularly in the domain of psychiatric classification [31]. In addition, variants of methodological parameters that may cause overfitting have been repeatedly discussed in prior review papers: sample-size limitation, in-sample validation, overhyping, data leakage, and especially "double-dipping" cross-validation (CV) methods [32, 33]. The cross-validation procedure is to evaluate the classification performance of the ML model by splitting the whole sample into an independent training set and testing set [32, 34]. Nevertheless, improper CV schemes have been found to overestimate model performance by "double-dipping" dependence or data leakage, which is a main source of incurring overfitting [8]. Besides, a recent review on the application of machine learning for gaining neurobiological and nosological insights in psychiatry underscored the need for cautious interpretation of accuracy in machine learning models [35]. That is, the analytic procedures to obtain reliable model performance are even more critical. However, a comprehensive review that systematically determines these methodological issues in prior studies of psychiatric machine learning classification is currently lacking, and how data/model availability allows for replication analysis to ensure generalization remains unclear. Thus, conducting a meta-research review concerning this topic would facilitate the characterization of the shortcomings and limitations in these current models. Moreover, developing a proof-of-concept assessment tool integrating these issues would facilitate the establishment of a favorable psychiatric machine learning eco-system.

To systematically access the generalizability issues, we conducted a pre-registered meta-research review of current studies that applied neuroimaging-based machine learning models to diagnose psychiatric populations. A total of 476 studies screened from PubMed (n $_{total}$=41,980) over the recent three decades (Jan 1990–July 2021) were included (see Additional file 1: Fig. S1-S2). First, geospatial mapping of the distribution patterns of the samples used in prior literature was depicted to illustrate the sampling biases. Furthermore, capitalizing on the sampling Gini model with the Dagum-Gini algorithm, we quantified the global and area-wide sampling inequality by taking both sampling biases and geospatial patterns into account. The underlying factors of these sampling inequalities were further explored, focusing on economic, social developmental, educational developmental gaps, and psychiatric disorder burdens, with a generalized additive model (GAM). Next, we focused on issues of poor generalizability by extending our examination to methodological issues that caused overfitting in previous psychiatric machine learning studies, which facilitated to

uncover potential pitfalls that may undermine generalizability. Finally, we utilized the results of our meta-research review to propose a 5-star standardized rating system for assessing psychiatric machine learning quality considering five domains: sample representativeness, cross-validation method, validation scheme for generalization assessment, report transparency, and data/model availability. Associations of study quality scores with publication year, psychiatric category, and model performance were then established.

Methods
The proposal and protocol for the current study have been pre-registered at Open Science Framework to endorse transparency.

Search strategy for literature
We searched eligible literature in accordance with PRISMA 2020 statement (Preferred Reporting Items for Systematic reviews and Meta-Analyses, see Additional file 1: Fig. S2). We retrieved literature at the PubMed database, with the following predefined criterion: (1) published from 1990 to 2021 (Jul); (2) peer-reviewed English-written article in journals or in conferences; (3) building machine learning models for diagnosis (classification) towards psychiatric disorders with neuroimaging-based biomarkers. By using Boolean codes and DSM classifications, we retrieved a total of 41,980 records from forty-eight 2nd level psychiatric categories. All records were input into Endnote X9 software for initial inspection and further underwent duplicate removal by using self-made code in Excel suits. Eligible papers were screened strictly following the inclusion and exclusion criteria detailed underneath. Furthermore, to obviate missing eligible records, we hand-inspected the reference list for the newest articles (2021).

We implemented a three-stage validation to ensure the correctness of all the processes. Stage 1: one reviewer was required to perform all the works (e.g., literature searching, data extraction, and data coding) by standard pipeline. Stage 2: a completely independent reviewer was asked to conduct all the works mentioned above for cross-check validation. Stage 3: another independent senior reviewer was designated to check the disparities of results between Stage 1 and Stage 2. If there were incongruences in records, the third reviewers should redo this process independently to determine which one was correct.

Inclusion and exclusion
We included studies by the following criteria: (1) machine learning models were built to diagnose (classify) psychiatric patients (defined by DSM-5) from healthy control

Chen *et al. BMC Medicine*   (2023) 21:241

Page 4 of 29

by neuroimaging-based biomarkers; (2) the ground-truth definition for patients was in accordance with clinical diagnoses performed by qualified staffs (e.g., clinical psychiatrists, DSM-5 or ICD-10); (3) fundamental information was given, such as bibliometric information, classifier, model performance, and sample size for both the training set and testing set. More details can be found in Additional file 1: Fig. S1.

We excluded studies that provided no original machine learning models and non-peer-reviewed results, including reviews, abstract reports, meta-analyses, perspectives, comments, and pre-printing papers. Furthermore, studies would be ruled out if they build models by non-machine learning algorithms or reported model performance with non-quantitative metrics. In addition, researches training machine learning models by non-neuroimaging-based features (e.g., genetics and blood markers) or in nonhuman participants were excluded in the current study. As aforementioned, we also discarded eligible studies if the patients' group had not yet been diagnosed by qualified institutes or medical staff. Finally, studies aimed at non-diagnostic prediction (e.g., prognostic prediction and regressive prediction) were removed for formal analysis.

Data extraction and coding
To ensure transparency and reproducibility, we extracted and coded data by referring to guidelines, including PRISMA [36], CHARMS checklist [37] (CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modeling Studies), and TRIPOD [38] (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) statement. As mentioned above, the three-stage validation was adopted here to ensure the correctness of these data. We coded eligible studies from two parts, with one for metainformation (e.g., publication year, affiliation, and countries for first author and journals) and another one for the scientific contexts of machine learning models (e.g., sample population, model performance, toolkit, feature selection methods, data availability, and sample size). Full contexts on data extraction and coding can be found in Additional file 1: Fig. S1.

Data resources
Less (more) economic developed countries (LEDC and MEDC) were defined by using the United Nations Development Programme (UNDP) criteria and International Monetary Fund (IMF, 2020) classification [39, 40]. Following that, a total of 34 countries or regions have been classified as MEDC, such as the USA, Germany, the UK, Japan, and Korea. Data for national development metrics derived from World Bank (WB)-World Development Indicators (2021), including Gross Domestic Product (GDP), Human Development Index (HDI), total government expenditure on public education (GEE), and research and development expenditure (R & D). In addition, we extracted data recording mental health disease burden (MHDB) and prevalence of psychiatric diseases from the Global Burden of Disease Study 2019 (GBD 2019) and the Global Health Data Exchange (GHDx) database. Finally, we obtained metrics for evaluating journal impacts by Journal Citation Reports of Clarivate ™ (2020).

Geospatial models
We built a global geospatial distribution model by packages of R, including the "ggplot2" and "maptools". The global geospatial map was defined by 251 countries or regions, which was validated by EasyShu suits. Furthermore, the geospatial maps for the USA, Germany, and the UK have been built by public dataset (CSDN communities). In addition, the geospatial pattern of China was built by the EasyShu software 3.2.2 for interactive visualization. Given the overlapping dataset, the global map visualizing the results of this geospatial model in the present study may be highly similar (but not equal) to the one in our previous work [22].

Sampling inequality coefficient
To quantify sampling bias and geospatial pattern for sampled population, we estimated sampling inequality based on the Dagum-Gini algorithm [41]. We estimated the Gini coefficient with Dagum-Gini algorithm by fitting multiple Lorenz curves, with absolutely high values for high sampling inequality. Specifically, we defined a relatively total sample size into each grid cell (e.g., each state in a country or each country in the world) based on extracted data in these eligible studies. Furthermore, the sub-modules were set by economic classification (i.e., MEDC and LEDC). Lastly, the Dagum-Gini model was used to decompose contribution from module-between variance, module-between-net variance, and intensity of transvariation. In this vein, we could estimate the Gini coefficient by adjusting the geospatial pattern and relative economic gap for a given economic entity, which improved statistical rigors by controlling unexpected variances. To validate the robustness of the Gini coefficient, we also calculated the Theil index based on the information entropy algorithm.

Case–control skewness
We calculated case–control skewness to estimate the extent to which the sample size between patients and the healthy control (HC) group was unbalanced, with a high value for high case–control skewness. We estimated the

Chen *et al. BMC Medicine*   (2023) 21:241

Page 5 of 29

ratio of the number of patients to HC when the sample size in the patient group was larger than the HC group and vice versa, which was used as a metric to quantify the case–control skewness.

Statistics

To examine the monotonic increasing trends for time-series data, we capitalized on the non-parametric Mann–Kendall Trend by using the R package [42]. Furthermore, we built both ARIMA (autoregressive integrated moving average) model and LSTM (long short-term memory) model to perform time-series prediction for the incremental trends of the number of relevant studies during the future decade, which were implemented by Deep Learning Toolbox embedded in MATLAB 2020b (Math-Works ® Inc.). Both models were trained by data split from 90% in the whole dataset and were tested in the remaining 10% dataset. Notably, we tested this model with real-world data using the actual number of relevant studies at the end of 2021 (Dec. 30) (see Fig. 1b).

Given the failure in fulfilling the prerequisites of parametric estimation, the Spearman rank model was defaulted for correlation analysis in the current study. Also, the parametric models for validating these

correlations have been built as well. Furthermore, the 95% confidence interval (CI) has been estimated by using Bootstrapping process at $n = 1,000$. Equivalent Bayesian analytic models have been constructed as well for providing additional statistical evidence. We used the Jeffreys-Zellner-Siow Bayes factor (BF) with prior Cauchy distribution ($r = 0.34$), with BF > 3 for strong evidence. To examine the non-linear associations of these variables of interest, we have built the generalized additive model (GAM) with natural shape-free spline functions by R package ("mgcv"). To obviate overfitting, the shape-free splines (i.e., smooth function) were used in these models. Finally, metrics of model performance (i.e., classification accuracy) for each study were precision-weighted rather than the original ones as reported.

Checklist for quantitative assessment on quality

We evaluated study quality in terms of the following five facets that were integrated from these meta-analytic findings: sampling representativeness (item 1: sample size and sites), model performance estimation (item 2: CV scheme), model generalizability (item 3: external validation), reporting transparency (item 4: reports for model performance) and model reproducibility (item 5: data/
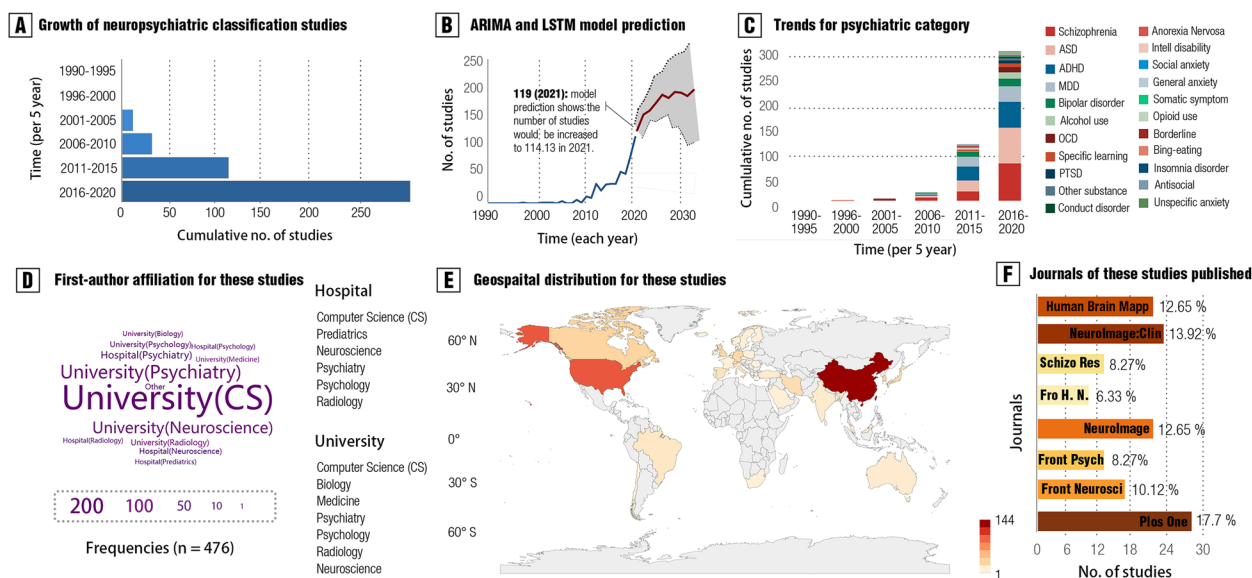

Fig. 1 Trends for research aiming at neuropsychiatric diagnostic prediction (classification) during the recent three decades (1990–2020). **A** illustrates the growth of the number of studies concerning neuropsychiatric classification from 1995 to 2020. **B** shows a prediction of the number of relevant studies for future decades based on both the autoregressive integrated moving average (ARIMA) model and the long short-term memory (LSTM) model. The number of relevant studies in 2021 was used as a testing set in the real world. We trained these models with data from 1990 to 2020 and tested them by using real data in 2021 to show the well generalizability. The models predicted the number of relevant studies would be increased to 114.13, and we found that the actual number of these publications in 2021 was 119. **C** presents trends for each psychiatric category during 1990–2021 (June). **D** shows the frequencies of first-author affiliation for all the included studies. **E** mapped the number of countries for the first affiliation in these included studies by using R packages "maptools" and "ggplot2". **F** illustrates which journals prefer to publish these studies. The top–bottom rank for these journals was determined by the number of these studies adjusted by the total number of publications per year. The length of the bar shows the proportion of one journal including these studies on all the journals

Chen *et al. BMC Medicine*   (2023) 21:241

Page 6 of 29

model availability). By ExpertScape™ rank and peer recommendations, we attempted to reach out to peer experts in multidisciplinary domains to examine the validity of this checklist, including data/computer science, psychiatry, neuroscience, psychology, clinical science, and open science. To this end, we have received concerns or advice for item classification and scoring criteria from three independent experts, and have performed four rounds of revision to form a final 5-star rating system called "Neuroimaging-based Machine Learning Model Assessment Checklist for Psychiatry" (N-ML-MAP-P). Three-stage validation was used to ensure assessment quality as well. Scores for one study would be reevaluated by a third independent scorer once the absolute difference between two scorers was larger than 2 points.

Results

General information

Four hundred and seventy-six studies with 118,137 participants from the 41,980 papers were eligible in this meta-research review (see Methods). These studies covered 66.67% (14/21) psychiatric disorders defined by the DSM-5 classification [43]. Diagnostic machine learning classifiers were mostly for schizophrenia (SZ, 24.57%, 117/476), autism spectrum disorder (ASD, 20.79%, 99/476), and attention deficit/hyperactivity disorder (ADHD, 17.85%, 85/476). To probe whether such research interests converged with the healthcare needs, we examined the association between the total number of machine learning studies concerning each psychiatric disorder and their prevalence/disease burdens (Data source: Global Health Data Exchange, GHDx) [44, 45]. Our findings indicated that there was no significant association between the number of studies related to different psychiatric disorders and their real-world prevalence (rho $= -0.24$, $p = 0.47$; $BF_{01} = 1.4$, moderate evidence strength for supporting the null hypothesis). Furthermore, we found no significant association between the number of these studies for psychiatric disorders and corresponding DALYs (i.e., disability-adjusted life-years)/ YLDs (i.e., years lived with disability) that reflected disease burden (DALYs, rho $= -0.05$, p $= 0.89$; $BF_{01} = 2.5$; YLDs, rho $= -0.06$, $p = 0.89$; $BF_{01} = 2.6$, moderate evidence strength for null association). Based on World Bank (WB) and International Monetary Fund (IMF 2021) classification, populations sampled in these studies were from 39 upper-middle-income to high-income countries, leaving population from the remaining 84.46% (212/251) countries in the globe unenrolled. In addition, 59.45% (283/476) of these studies used domestically-collected samples, while 31.10% (148/476) reused open-access datasets (e.g., ABIDE and ADHD-200).

Historical trends

The total number of psychiatric machine learning studies for diagnostic classification on psychiatric disorders increased markedly in the past 30 years ($z = 5.81$, $p = 6.41 \times 10^{-9}$, Cohen $d = 1.82$, Mann-Kendell test) (see Fig. 1a and Additional file 1: Fig. S3). Based on time-series prediction models, we predicted a persistent increment for the number of studies pursuing brain imaging-based diagnostic classification for psychiatric disorders in the future decade (e.g., $k = 229.65$ in 2030, 95% CI: 106.85–352.44) (see Fig. 1b and the " Methods" section).

Despite the accelerated increase in the number of psychiatric machine learning studies, the increase rate for different psychiatric disorders was found to be different: the number of existing studies on SZ, ASD, ADHD, major depression disorder (MDD), and bipolar disorder (BP) is significantly larger than that on other high-disease-burden categories (e.g., eating disorder and intellectual disability) (see Fig. 1c). To quantify the increment pattern for different psychiatric categories, we capitalized on increment curve models. We found that increase speeds for machine learning models regarding neuropsychiatric diagnoses towards SZ ($b = 2.40$, 95% CI: 2.05–2.74, $p < 0.01$) and ASD ($b = 2.64$, 95% CI: 2.25–3.02, $p < 0.01$) were significantly faster than others (see Additional file 1: Fig. S3 and Tab. S1).

Interestingly, a quite number of first authors of these studies (46.42%, 221/476) seemed to be trained in computer and data science instead of psychiatry or neuroscience (see Fig. 1d). Institutes from China, the USA, Canada, Korea, and the UK contributed mostly for the total number of these machine learning studies (see Fig. 1e). Moreover, by adjusting the total publications per year, we found that these studies were mostly published in journals with a special scope on neuroimaging, such as *Human Brain Mapping* and *Neuroimaging: Clinical* (see Fig. 1f and Additional file 1: Tab. S2-S3).

Sampling bias and sampling inequality
Geospatial pattern of sampling bias

Geospatial maps were generated to visualize the distribution of the sampled populations (i.e., the number of participants). We found that the sample populations covered only the minority upper-middle-income and high-income countries (UHIC) worldwide ($n_{\text{UHIC countries}} = 32$; 12.74%). Even in UHIC, across-country imbalance in sample population was striking (total sample size: $n_{\text{Chinese}} = 14,869$, $n_{\text{Americans}} = 12,024$, $n_{\text{Germans}} = 4,330$; see Fig. 2a and Additional file 1: Tab. S4). Moreover, we found a likewise prominent within-country imbalance of sample populations (see Fig. 2b, Additional file 1: Tab. S5-S8 and Additional file 1: Fig. S4-S5). Furthermore, as

Chen *et al. BMC Medicine*   (2023) 21:241

Page 7 of 29

A **Sampling population for ML in the globe (1990-2020)**
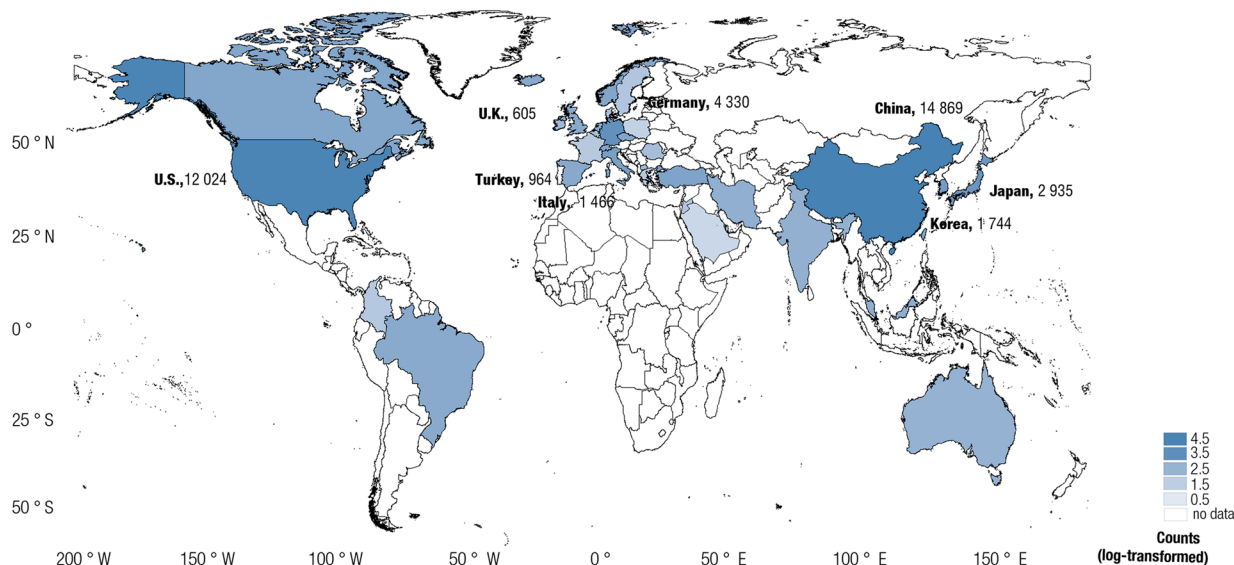
B **Sampling population for ML within the U.S. (1990-2020)**
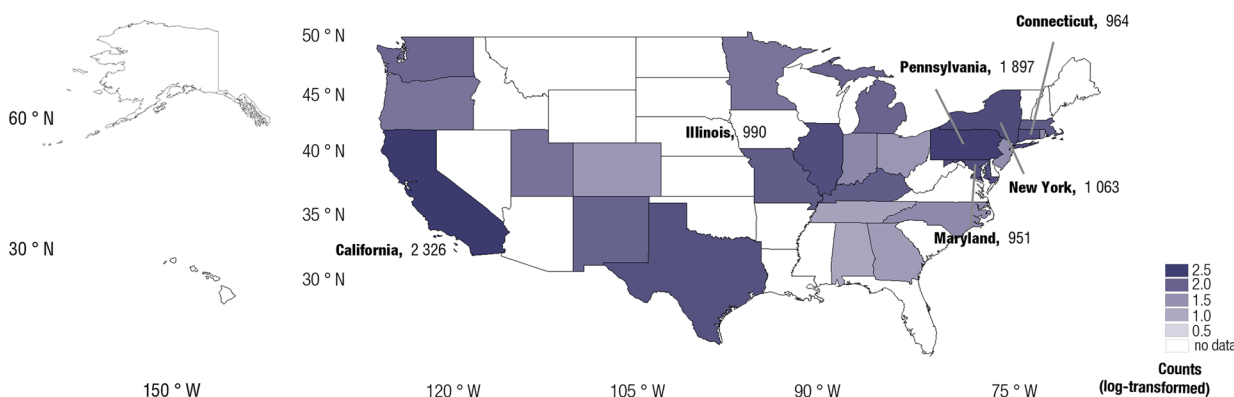

Fig. 2 Geospatial model for sample population regarding ML models towards neuropsychiatric classification in the world (**A**) and USA (**B**). Both maps were built by 1st administrative grid cell, with each country/region for the globe (251 countries/regions) and state for the USA (51 states). For better readability, we re-scaled the sample size by log-transformation. Sample size for a portion of countries/regions has been shown in these maps. **A** panel was depicted similarly to the Fig. 2A in our previous article [22], because of the overlapping datasets between them

for continents-based classification, populations of these machine learning models were largely enrolled from Asia (44.67%, adjusted by total population) and North America (26.76%, adjusted by total population). Notably, in the current meta-research review, no machine learning models were observed to train classifiers by samples in Africa despite its large population.

We examined whether the size of the sample population in these models could be determined by the national economic level. Results showed a strikingly positive association between nation-wide GDP (Gross Domestic Product, Data source: IMF 2021) and total sample size all over the globe ($r = 0.65$, 95% CI: 0.40–0.81, conditions-adjusted, $p < 0.001$; $BF_{10} = 1.10 \times 10^3$, Strong evidence) (see Fig. 3a). Supporting that, such association was found within China ($r = 0.47$, 95% CI: 0.02–0.76, $p < 0.05$, conditions-adjusted; $BF_{10} = 1.93$, moderate evidence) and the USA ($r = 0.47$, 95% CI: 0.10–0.73, $p < 0.05$, conditions-adjusted; $BF_{10} = 3.72$, Strong evidence), respectively (see Fig. 3b–c).

Sampling inequality
To quantitatively evaluate such sampling bias, the new concept, sampling inequality, was introduced, which
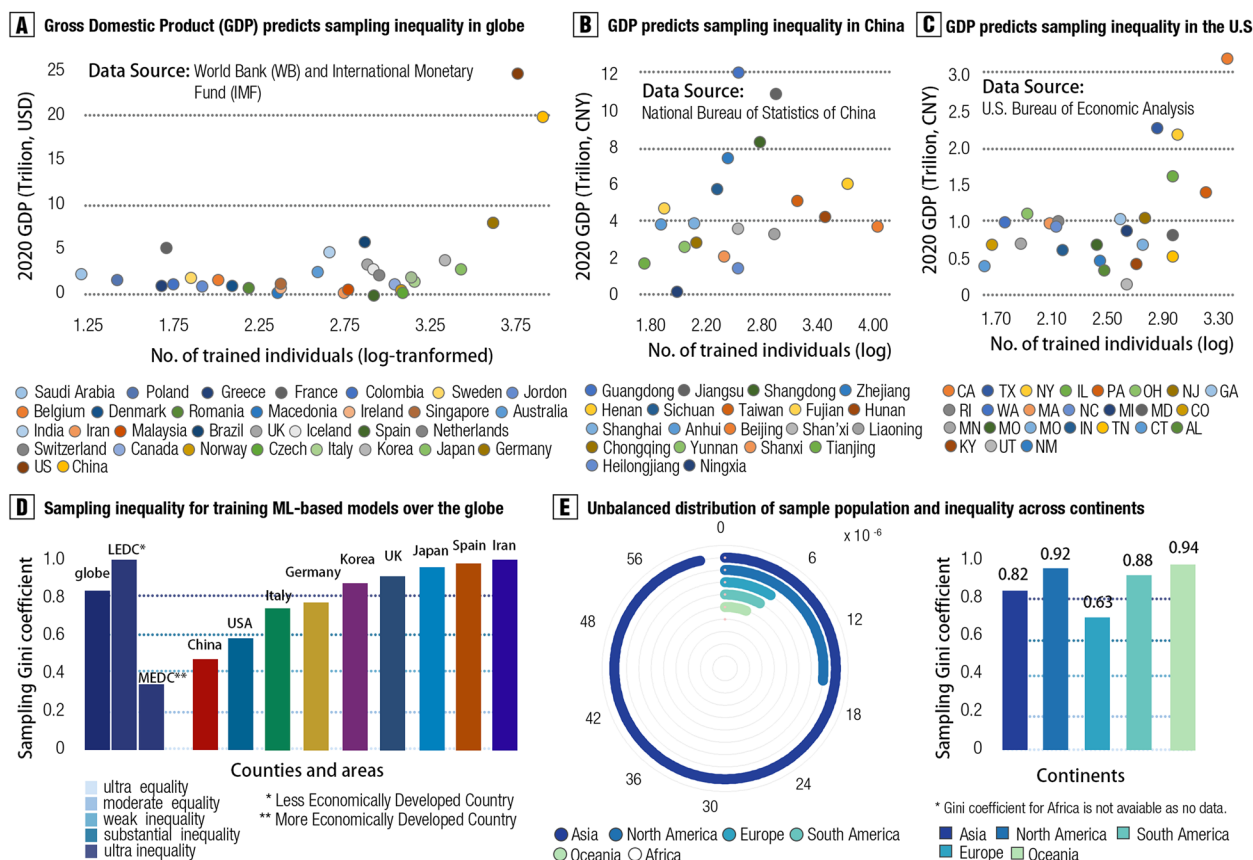
Fig. 3 Sampling bias and sampling inequalities in these trained ML models. **A** provides a scatter plot for the association between GDP and sample size for 32 counties/regions in the globe. **B** offers a scatter plot showing the association between GDP and sample size for 20 provinces within China. (C) shows the association between GDP and sample size for 25 states within the USA. (D) plots Gini sampling coefficients for the top 10% countries with large sample sizes to train ML models in existing studies, with high Gini value for high sampling inequality. LEDC and MEDC were categorized by World Bank (WB) and International Monetary Fund (IMF) classification. **E** illustrates the sampling bias and Gini coefficients for each continent. The left panel shows the proportion of the total sample size for training ML models in existing studies on the total sample population for each continent. The right panel shows the Gini coefficient for each continent

reflects both the sample-size gap and the geospatial bias for the sampled populations reported in existing psychiatric machine learning studies. We used sampling Gini coefficient (*G*, ranged from 0 to 1.0) based on the Dagum-Gini algorithm, to quantify the degree of sampling bias (see Methods). We found severe sampling inequality in samples of prior psychiatric machine learning studies ($G = 0.81$, $p < 0.01$, permutation test, see Fig. 3d). Furthermore, based on IMF classification, we grouped global countries into More Economically Developed Country (MEDC) bloc and Less Economically Developed Country (LEDC) bloc and found a significant difference in the sampling inequality between them: sampling Gini coefficient in LEDC was threefold ($G_{LEDC} = 0.94$, $G_{MEDC} = 0.33$, $p < 0.01$, permutation test) higher than that in MEDC. In addition, we also examined within-country sampling inequality. Results showed a weak sampling inequality in China ($G = 0.47$) and the USA

($G = 0.58$), but severe inequality in Germany ($G = 0.78$), the UK ($G = 0.87$), Spain ($G = 0.91$), and Iran ($G = 0.92$) (see Fig. 3d and Additional file 1: Tab. S9). Furthermore, we found a relatively lower sampling inequality in Europe compared with other continents ($G_{Europe} = 0.63$; see Fig. 3e and Additional file 1: Tab. S10-S11). Notably, a significantly positive association between these sampling Gini coefficients and averaged classification accuracy was uncovered ($r = 0.60$, $p = 0.04$, one-tailed; permutation test at $n = 10,000$), which possibly implied potential inflated estimates for model performance because of such sampling inequality.

To examine whether sampling inequality was further increased by economic gap, that was, individuals (patients) living in richer countries (areas) were more likely to be recruited in building rich-areas-machine learning-specific models, a generalized additive model (GAM) with natural shape-free spline function was

constructed. Interestingly, the GDP of these countries allows for an accurate prediction of the sampling inequality values ($\beta = -2.75$, S.E $= 0.85$, $t = 4.75$, $p < 0.001$, $R^2_{adj} = 0.40$; $r = -0.84$, 95% CI: -0.41 to -0.97, $p < 0.01$; $BF_{10} = 13.57$, strong posterior evidence), with higher national income for weaker sampling inequality. The apparent presence of sampling bias and high sampling economic inequality for the reviewed psychiatric machine learning studies may resonate with generalization failure that was widely concerned in the field.

Methodological considerations on generalizability
Sample size, validation, technical shifting, and case–control skewness

We extended the investigations of sampling bias and sampling inequality to an analysis of other methodological facets that may likely lead to overfitting and hence magnify the generalization errors. A significant correlation was found between the sample size in psychiatric machine learning studies and publication year in the last three decades ($r_{(total)} = 0.75$, 95% CI: 0.22–0.93, $p = 0.013$; $BF_{10} = 5.83$, Strong evidence) (see Fig. 4a and Additional file 1: Tab. S12-S14). Despite improvement over time, we observed a strikingly biased distribution skewing to a small sample size ($n < 200$) in these machine learning models (73.10%, 348/476) (see Fig. 4b and Additional file 1: Tab. S15).

In addition, we found a prominently positive association between the ratio of using k-fold cross-validation (CV) scheme and publication year in recent decades ($r = 0.82$, 95% CI: 0.40–0.95 $p < 0.01$; $BF_{10} = 15.80$, Strong evidence) (see Additional file 1: Tab. S16-S17). As repeated recommendations by didactic technical papers [8, 10, 34, 46], adopting a k-fold CV to validate model performance could outperform popular LOOCV methods in terms of model variance and biases. We thus examined model performance between them by precision-weighted method [47] that could adjust the effects of sample size and between-study heterogeneity. Results showed that model performance estimated by
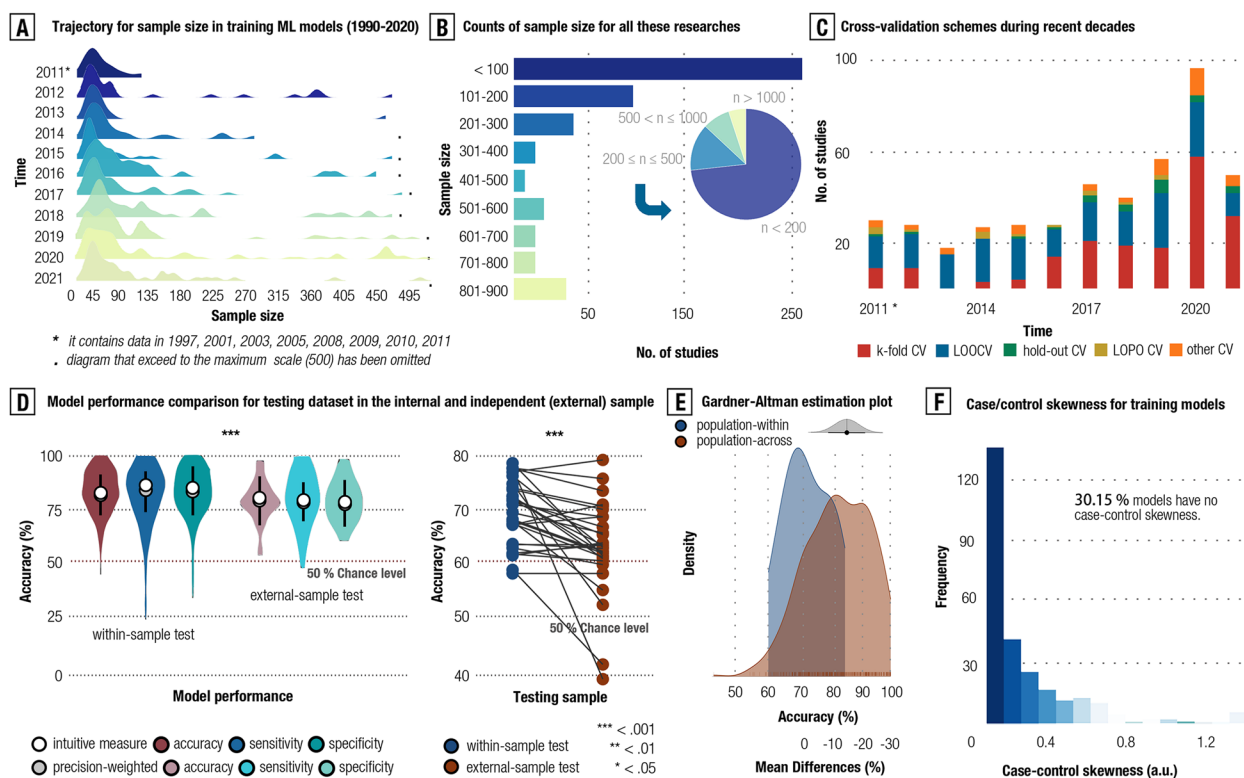

Fig. 4 Methodological considerations for existing ML models towards psychiatric diagnosis. **A** illustrates increment trends for sample size during the recent three decades by Gaussian kernel density plots. Labeling 2011 sums up all the sample size from 1990 to 2011. **B** shows the counts for subgroups by dividing these studies according to sample size. **C** plots the trends of using cross-validation (CV) schemes by accounting counts from all the included studies during the recent three decades. **D** shows model performance comparisons between independent-sample validation and within-sample validation. The non-parametric W test was used for statistical inferences, with *** for $p < .00$. Precision-weighted accuracy was estimated by Woo et al. **E** depicts Gardner-Altman estimation for the classification accuracy comparison between population-within sample and population-across sample. Black dot indicated the point estimate for the mean difference (delta) of the two groups, and the shadow areas showed the distribution estimated by delta. **F** presents a frequency plot to show case–control skewness

LOOCV was prominently higher than k-fold CV (Acc $_{LOOCV}$ = 80.35%, Acc $_{k-fold}$ = 76.66%, precision-adjusted, w = 20,752, p < 0.001, Cohen d = 0.31; BF$_{10}$ = 2289, strong evidence) (see Fig. 4c). Details for other methodological considerations can be found at Additional file 1: Tab. S18-22.

As for independent-sample validation, we found a significantly positive association between the ratio of validating model performance in the independent sample (site) and publication year in recent decades (r = 0.88, 95% CI: 0.63–0.97 p < 0.01; BF$_{10}$ = 234.93, Strong evidence). Nevertheless, the majority of these machine learning studies (84.24%, 401/476) still lacked validation for model generalizability in the independent sample(s). Furthermore, we found that the classification performance of these models tested in the independent samples was more "conservative" than those tested in the internal samples (Acc $_{independent-sample\ validation}$ = 72.71%, Acc $_{others}$ = 77.75%, precision-adjusted, w = 3,041, p < 0.001, Cohen d = 0.32; BF$_{10}$ = 29.43, Strong evidence) (see Fig. 4d and Additional file 1: Tab. S23). To directly test the impact of sampling bias on model generalizability, we compared the model performance between cross-country samples (i.e., training model in a sample from one country and testing model in a sample from other countries) and within-country (i.e., training and testing model in sample within the same countries) sample. Results showed that model performance was more "conservative" in the cross-country sample than in the within-country sample (Acc $_{cross-country\ sample}$ = 72.83%, Acc $_{within-control\ sample}$ = 82.69%, precision-adjusted, w = 2,008, p < 0.001, Cohen d = 0.54; BF$_{10}$ = 150.90, Strong evidence; see Fig. 4e).

Furthermore, we specifically examined the shift of mainstream neuroimaging modalities and features of these models in recent decades. Results showed that the (functional) MRI was still the mainstream neuroimaging technique to build these models over the last three decades (i.e., averaged 73.70% of these models for (functional) MRI, 21.57% for EEG/ERP, 2.69% for fNIRs, 2.02% for MEG and 0.21% for PET). Despite that, the increasing trend of using multi-modalities in training these neuroimaging-based ML models was observed, from 3.22% to 19.19% of these models over time. In addition, with the developments of ML techniques, the ratio of using deep learning models or complicated parameterized models to "shallow learning models" was increasing during recent decades, particularly after 2019 (i.e., 0% in 2012, 10.40% in 2016, and 32.32% in 2020). As for the strategy of feature selection, we found an increase in the applications of algorithmic techniques than of pre-engineered selections in building these models (i.e., 0% before 2012, and averaged 30.90% after 2012). Nevertheless, no changes were found for the shift of paradigm from a single-snapshot case–control cohort to repetitive scanning of the same participants in these models. While the shifting of main neuroimaging modalities, model complexity, and feature selection strategy was observed over time, we found no prominent trends of model performance (i.e., precision-weighted accuracy) over time (Accuracy: 84.43%, 95% CI: 81.84–87.88% at 2011; 84.38%, 95% CI: 80.79–87.86% at 2015; 84.78%, 95% CI: 82.82–87.49% at 2020). Full results for these findings can be found in Additional file 1: Fig. S6-S8.

Finally, by calculating the standardized case–control ratio (see the "Methods" section), we observed a case–control skewness (i.e., the number of patients is larger than healthy control, and vice versa) in a quarter of all the included studies (25.37%, 121/476) (see Fig. 4f). The case–control skewness was significantly (but weakly) associated with the reported classification accuracy, which may imply inflated accuracy due to the imbalanced case–control distribution in the data (r = 0.15, 95% CI: 0.04–0.27, p < 0.05; BF$_{10}$ = 2.04, moderate evidence).
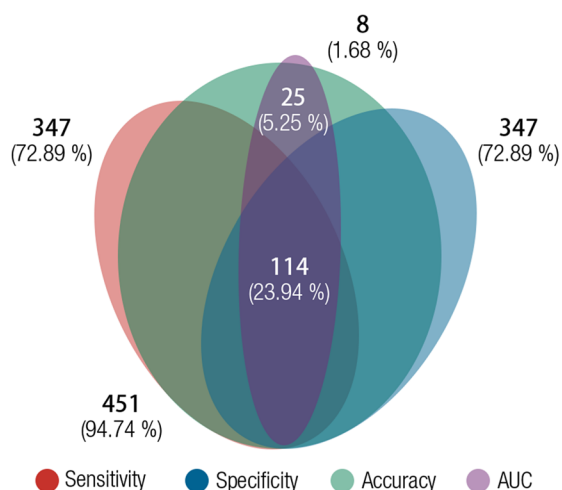
Technical transparency and reproducibility

We further determined whether existing studies provided sufficiently transparent reports to evaluate potential overfitting and reproducibility. We found that only one fifth of them (23.94%, 114/476) fulfilled the minimum requirements for reporting model results (i.e. balanced accuracy, sensitivity, specificity, and area under curve) by the criterion as proposed by Poldrack [8, 48] (see Fig. 5a).

As for the model reproducibility, only 12.25% (58/476) of studies shared trained classifiers (full-length codes). Furthermore, only 19.12% (91/476) studies claimed to provide available original data. Notably, we manually checked the validity of these resources as these studies stated, one-by-one, but found that only a small portion of trained classifiers (32.27%, 19/58) or data (15.38%, 14/91) were actually available/accessible (see Fig. 5b). Thus, incomplete reports for model results and poor technical reproducibility may be one of the sources to hamper the assessment of generalizability, and hence, the "generalization crisis" remains.

Five-star quality rating system

To promote the establishment of an unbiased, fair, and generalizable diagnostic model, we proposed a 5-star quality rating system called "Neuroimaging-based Machine Learning Model Assessments Checklist for Psychiatry (N-ML-MAP-P)" by integrating these meta-research findings aforementioned and up-to-date guidelines that provided by multidisciplinary experts (see Methods). This rating system incorporated five elements, including sample representativeness, CV methods,

Chen *et al. BMC Medicine*   (2023) 21:241

Page 11 of 29

A **Reporting transparency for these ML studies**   **B** **Technical availability for these ML studies**
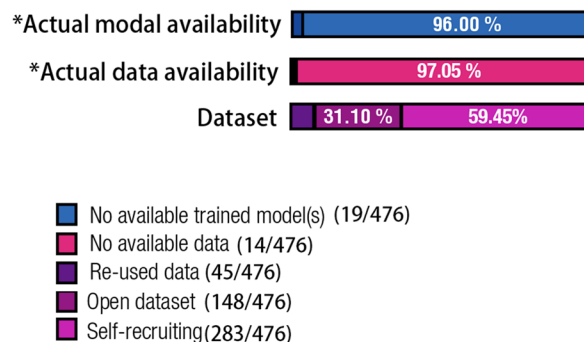

Fig. 5 Reporting transparency and technical (data and model) availability. **A** presents patterns of reporting model performance across sensitivity, specificity, balanced accuracy, and area under curve (AUC) by a Venn plot. **B** sums up the proportion of having actual model availability, data availability, and datasets

independent-sample validations, reports for model performance, and data/model availability (see Fig. 6a).

Based on this N-ML-MAP-P rating system, we found that overall quality scores for these models have increased consistently over the last decade ($r = 0.77$, 95% CI: 0.25–0.99, $p < 0.01$; $BF_{10} = 7.04$, strong evidence), demonstrating that study quality for machine learning models on psychiatric diagnosis has been increasingly improved (see Fig. 6b). In addition, we also examined study quality for each item and revealed that ratings for sample size, CV methods, independent validation, and reporting transparency have been gradually improved (see Additional file 1: Tab. S24). However, we found no prominent increase in quality scores on technical (data and model) availability (see Additional file 1: Tab. S25). Furthermore, we found a considerably strong positive correlation between the number of disorder-specific studies and their quality scores ($r = 0.69$, 95% CI: 0.13–0.88, $p < 0.05$; $BF_{10} = 4.10$, strong evidence), with relatively high quality for machine learning studies concerning SZ, ASD, and ADHD.

Despite the increase, the overall quality scores remained relatively low in the vast majority of these models (see Fig. 6c–d). Intriguingly, we found a weak but statistically significant association between journal impact factors/journal citation indicator (JIF/JCI) and the scores of model quality rated by N-ML-MAP-P assessment ($r_{(JIF)} = 0.18$, 95% CI: 0.08–0.30, $p < 0.001$; $BF_{10} = 41.90$, strong evidence; $r_{(JCI)} = 0.15$, 95% CI: 0.06–0.25, $p < 0.01$; $BF_{10} = 8.60$, strong evidence) (see Fig. 6e). Furthermore,

we also observed a weakly negative association between the JIF/JCI and model performance ($r = -0.19$, 95% CI: -0.10 to -0.28, $p < 0.001$; $BF_{10} = 4697.67$, strong evidence) (see Fig. 6f).

In summary, our purpose-built N-ML-MAP-P system for quantitatively assessing the quality of these models revealed prominent improvements for them over time, possibly indicating that efforts made by scientific communities [8, 10, 49] to address overfitting issues in diagnostic machine learning models for psychiatric conditions may be effective. However, existing machine learning studies may still face several challenges, e.g., low overall quality and poor technical reproducibility, which still characterize a majority of these studies. A full list of these models can be found in Additional file 2 [50–510].
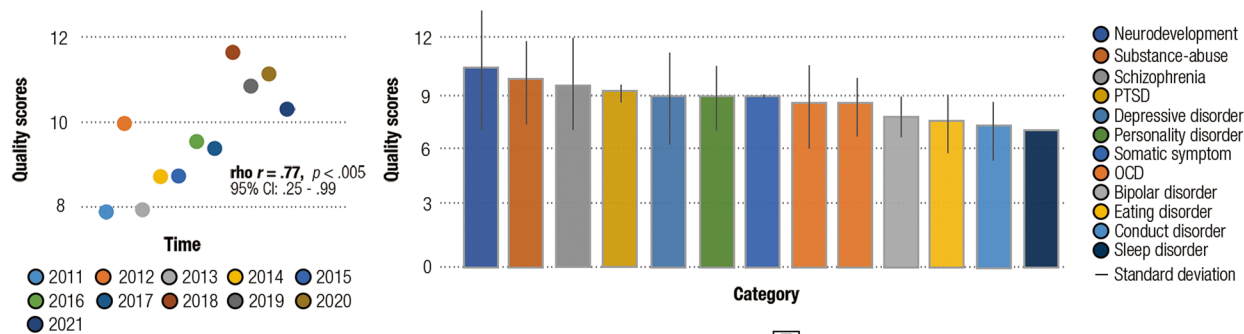
Discussion

We conducted a pre-registered meta-research review and quantitative appraisal to clarify generalizability and even quality in existing machine learning models on neuroimaging-based psychiatric diagnosis ($k = 476$) from insights into sampling issues, methodological flaws, and technical availability/transparency. By doing so, we quantified a severe sampling economic inequality in existing machine learning models. By further determining methodological issues, we found that sample-size limitation, improper CV methods, lack of independent-sample validation, and case–control skewness still contributed to an inflation of model performance. Furthermore, we found a poor technical availability/transparency which may in turn
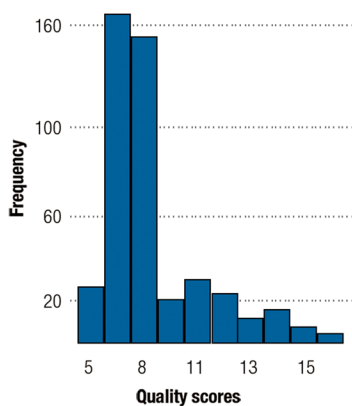
Fig. 6 Neuroimaging-based machine learning model assessment checklist for psychiatry (N-ML-MAP-P). **A** provides details for five items and scoring criteria in this checklist for evaluating the study quality of all the included studies. **B** presents a scatter plot for showing the trends of improving study quality during the recent decade (2011–2021). **C** shows the overall study quality for each psychiatric category in existing studies. This plot is ranked by total quality score, and bars indicate standard error (S.E.). **C** provides a frequency plot for overall quality scores. **D** shows the trajectories of study quality for different affiliations, including data/computer science, neuroscience, psychiatry, and others. **E** draws a scatter plot showing the association between journal quality (i.e., journal impact factor, JIC) and overall quality scores. **D** provides a scatter plot to show the association of overall quality scores with model accuracy as reported in these studies

critically hamper mechanisms to examine generalizability for these models. Based on these findings, we developed a checklist to quantitatively assess the quality of existing machine learning models. We found that despite increasing improvement, the overall quality of the vast majority of these machine learning models was still low (88.68%

models were rated at low quality in existing literature). Taken together, the results indicated that ameliorating sampling inequality and improving the model quality may facilitate to build of unbiased and generalizable classifiers in future clinical practices.

One critical finding that warrants further discussion is the severe global sampling inequality in existing machine learning models. Despite rapid proliferation, we found that the samples were predominately recruited from upper-middle-income and high-income countries (444/476 models, 93.28%). Also, we observed regional sampling inequalities, with the Gini coefficient in LEDC being 3-fold higher than that in MEDC. To make matters worse, both sampling inequalities were found to be enlarged by regional economic gaps. Despite supporting the descriptive observation to previous studies [22, 35, 511], the present study provided unique statistical evidence to clearly reveal the severity of sampling bias of these extant models in the globe or across countries (regions), which advanced our knowledge to make the sampling bias quantitatively comparable rather previously conceptional concern. Beyond that, the predictive role of national (regional) economic level on these sampling biases has been quantitatively verified, possibly indicating a sampling economic inequality in the neuroimaging-based ML-aid diagnosis. For instance, in China, machine learning models were predominately trained by samples from mid-eastern Chinese with high incomes, whereas there was no evidence to validate whether these trained machine learning models could be generalized for western Chinese with lower incomes. Notably, it is the same case with predictive models, recent works have revealed generalization failure for cross-ethnicity/race samples in neuroimaging-based predictive models [512, 513]. Compared to previous studies using qualitative inferences or conclusions [514, 515], we provided preliminary evidence to quantify the extent to which sampling inequality impacted model generalizability, this result may imply how much sampling inequality should be limited to built generalizable neuroimaging-based diagnostic classifiers. As a didactic example recently, Marek and colleagues (2022) have provided a quantitatively empirical evidence that thousands of subjects are needed to attain reliable brain-behavior associations, although the intensive discussions or concerns for the high overfitting in neuroimaging-based models with small sample sizes have been debated for a long-lasting time [10, 34, 46, 516]. Thus, by quantitatively revealing the association of sampling issues to inflated classification accuracy, the present study may provide valuable insights into how to increase sampling equality enough to achieve good model generalizability in future empirical studies. To tackle this issue, diversifying sample representation (i.e., racial balance or

socioeconomic balance) in neuroimaging-based predictive models has been increasingly advocated [517, 518]. That is, existing "population-specific models" trained with less or even no samples in LEDC and Africa practically questioned their generalizability across intersectional populations. More importantly, besides common sense for the disadvantages of global economic gaps in science development, "leaving the poor ones out" in training machine learning models may not only render poor generalizability to psychiatric diagnostic models but also exacerbate global inequalities in clinical healthcare. Relating to this consideration, future studies could explore whether and how economic gaps contribute to biases in clinically diagnostic measures, such as neuroimaging-based precision diagnosis in high-income countries, as opposed to more subjective symptom-dependent diagnosis in low-income countries. This investigation could further our understanding of how economic disparities impact the inequalities in the development and implementation of diagnostic basis. Nonetheless, another insightful viewpoint worthy to note was that an overly board emphasis on generalizability may impede clinical applications of these machine learning models in specific medical systems (e.g., healthcare) [519, 520], with high generalizability at the expense of optimal model performance within specific cohorts. In other words, despite poor geospatial or socioeconomic generalizability, these machine learning models posing high performance within specific contexts (e.g., machine learning model trained by data in the A hospital could accurately predict patients within A hospital rather than other ones) may be still reliable into a given clinical practice.

Another factor contributing to poor generalizability was rooted in methodological issues. We found that, with consistent efforts made by scientific communities [8, 521], the ratio of using k-fold CV in estimating model performance gradually increased during recent decades, which may partly mirror effective controls for overestimation on diagnostic accuracy that was caused by flawed CV scheme [23, 34]. However, the LOOCV was still used widely (40.33%) in recent decades, which may overfit models compared to those using k-fold CV (precision-weighted classification $Acc_{level-one-subject-one\ CV}$, 80.35%; $Acc_{k-fold\ CV}$, 76.66%, $p < 0.001$). Thus, although the repeated technical recommendations and calls may be effective in changing our practices to rectify model overfitting, this issue has not been fully addressed to date [522, 523]. Compared to the CV method, testing model performance in external (independent) samples could provide more accurate estimates for interpretability and generalizability [524, 525]. Nevertheless, only 15.76% models were validated in the independent sample (s). More importantly, we observed that model performance

may be highly overestimated in within-country independent samples compared to cross-country ones (precision-weighted classification Acc $_{cross-country}$, 72.83%; Acc $_{within-country}$, 82.69%, $p < 0.001$). Thus, not only "independent-sample" validation but also the well-established "intersectional-population-cross" validation is demanded to strengthen generalizability in future studies [526]. Moreover, few machine learning models (< 5%) provided adequate technical availability, which diluted our confidence for the generalizability and reproducibility of these models, especially in the "big data" era [527]. On balance, we found that methodological flaws of these machine learning models were increasingly ameliorated to prompt model generalizability in recent decades, but sample limitation, improper CV methods, lack of "cross-population" independent-sample validation, and poor technical availability still exposed these models to high risks of overfitting.

In the current study, we proposed a quantitative framework for evaluating the quality of these models, covering sample, CV, independent-sample validation, transparency, and technical availability. We found that the overall quality of these models increasingly improved over time. As aforementioned, some didactic methodological papers [8, 9, 23] have considerably contributed to prompting the scientific communities to rectify these methodological flaws in machine learning models. Furthermore, reporting benchmarks or guidelines were also developed to increase the transparency of information for accurately evaluating model performance in recent years [38, 528]. However, despite encouraging improvements, the low-quality machine learning models seem to still dominate this field, as we observed in the current study that single-site samples and poor data/model availability remained largely unchanged [15]. Together, the findings may imply that existing machine learning models are not as solid as claimed in terms of generalizability and reproducibility in clinical practices in their current form. It is noteworthy that the high-quality models that were rated in the current study have not yet been tested for generalizability. Thus, testing the generalizability or reproducibility of these models from originally trained samples to different populations (e.g., countries, ethnics, income-levels) could be a more reliable and valid way to validate the generalizability in future studies.

To tackle these generalizability issues, here we recommended several practical tips. Beyond sample size, recruiting a diverse, economically-equal, case–control balanced, and representative sample is one of the best avenues to obviate sampling biases. Technically speaking, the cross-ethnicity/race or cross-country independent sample should be prepared for the generalizability test. At least, the nest k-fold CV method

is clearly warranted. Moreover, we also recommend transparent and unfolded reports for model performance facilitating to take these models into clinical insights. In addition, improving the data/modal availability for these models is one of the ways to provide venues to validate clinical applicability. To conceptualize and streamline these recommendations, we have preliminary built the "Reporting guideline for neuroimaging-based machine learning studies for psychiatry" (RNIMP 2020) checklist and diagram (see Additional file 3: Tab. S1-2); this suit is developed by encapsulating the above tips and currently promising benchmarks [8, 23].

This study warrants several limitations. First, we narrowed the research scope into diagnosis but not all the categories were sampled evenly. Predictive machine learning models for psychiatric conditions included (at least) three forms of prediction: diagnosis (i.e., predicting the current psychiatric condition), prognosis (i.e., predicting outcome in future onsets), and prediction (i.e., predicting response or outcome for a given treatment) [529]. Second, this study may not cover all eligible data, especially in literature that was published in African areas or written in non-English languages, as data were screened from English-written peer-reviewed papers. Therefore, we stress that all the findings are grounded on these studies, instead of completely representing the real-world situation. Third, the current study did not probe into the model generalizability from biological insights, but focused on sampling bias and methodological issues only. Thus, it left room to be uncovered in future works. Fourth, we empirically inferred the academic training experiences of the first authors by their affiliation. However, such assumptions may not be solid. Extending these conclusions from this section to elsewhere should be more prudent. Fifthly, the present study has not thoroughly analyzed the factors that contribute to the changes in the methodological and technical underpinnings of machine learning models for psychiatric diagnosis. For instance, the increased sample size in these models over the last decades may be attributed to the improvements of imaging techniques/infrastructures, the developments of machine learning knowledge, and the decrease of data costs, which are not explored in the current study. In other words, future studies could reap huge fruits from delving into the specific roles of these factors in the advances of these models, particularly in sample size. Lastly, given that the neuroimaging-based signatures are not practically applicable for diagnosing all psychiatric conditions, the statistics for unbalanced developments and qualities across these DSM categories should be explained more prudently.

Conclusions

On balance, we provided meta-research evidence to quantitatively verify the sampling economic inequality in existing machine learning models for psychiatric diagnosis. Such biases may incur poor generalizability that impedes their clinical translations. Furthermore, we found that the methodological flaws have been increasingly ameliorated because of repeated efforts made by these technical papers and recommendations. Nonetheless, in the present study, we stretched views to find that these limitations including small sample size, flawed CV method (i.e., LOOCV), no independent-sample validation, case–control skewness, and poor technical availability still remained, and have demonstrated quantitative associations of such limitations to inflated model performance, which may hence indicate model overfitting. In addition, poor reporting transparency and technical availability were also observed as a hurdle to translate these models into real-world clinical actions. Finally, we extended to develop a 5-star rating system to provide a purpose-built and quantitative quality assessment of existing machine learning models and found that the overall quality of a vast majority of them may still be low. In conclusion, while these models showed a promising direction and well-established contributions in this field, it is suggested that enhancing sampling equality, methodological rigor, and technical availability/reproducibility may be helpful to build an unbiased, fair, and generalizable classifier in neuroimaging-based machine learning-aid diagnostics of psychiatric conditions.

Abbreviations

Acc	Accuracy
ADHD	Attention deficit/hyperactivity disorder
ARIMA	Autoregressive integrated moving average
ASD	Autism spectrum disorder
BP	Bipolar disorder
CV	Cross-validation
DALYs	Disability-adjusted life-years
GAM	Generalized additive model
GBD	Global Burden of Disease
GDP	Gross Domestic Product
GEE	Total Government Expenditure on public Education
HC	Healthy control
HDI	Human Development Index
LEDC (MEDC)	Less (more) economic developed countries
LOOCV	Leave-one-out cross-validation
LSTM	Long short-term memory
MDD	Major depressive disorder
MHDB	Mental health disease burden
ML	Machine learning
OSF	Open Science Framework
R & D	Research and Development expenditure
ROB	Risk of bias
SZ	Schizophrenia
UHIC	High-income countries
WB	World Bank
WEIRD	Western, Educated, Industrialized, Rich, and Democratic
YLDs	Years lived with disability

Supplementary Information

Additional file 1: Figure S1-S8 and Table S1-S25. **FigS1.** Research pipelines for data acquisition. **FigS2.** PRISMA 2020 flow diagram for the current study. **FigS3.** Trends in ML-based diagnostic prediction for psychiatric diseases by neural features. **FigS4.** Mental health disorders as the portion of total disease burden at 2019. **FigS5.** Geospatial model for sampling population within China, Germanyand U.K. **FigS6.** Distribution of methodological details. **FigS7.** Model performance across algorithm, tookit, cross-validation, sample sizeand skewness. **FigS8.** Model performance across validations, trajectories, psychiatric categories, journal impacts, scanning technology/modalityand institutes/datasets. **TabS1.** Curve fitting results for exponential function model. **TabS2.** Journals counts for papers aiming at neuropsychiatric diagnostic prediction. **TabS3.** Counts for contributors' sources for these papers. **TabS4.** Summary for sample population for these papers in the world. **TabS5.** Summary for sample population for these papers in the U.S. **TabS6.** Summary for sample population for these papers in the China. **TabS7.** Summary for sample population for these papers in the Germany. **TabS8.** Summary for sample population for these papers in the U.K. **TabS9.** Sampling inequalities for globe and countries/regions. **TabS10.** Sampling inequalities for continents. **TabS11.** Sampling inequalities and national development index. **TabS12.** Sample size during recent decade for all the studies. **TabS13.** Sample size during recent decade for studies using self-recruiting sample. **TabS14.** Sample size during recent decade for studies using open dataset. **TabS15.** Sample size during recent three decades in the current study. **TabS16.** Summary for what modelswere built for neuropsychiatric diagnostic prediction in existing studies. **TabS17.** Summary for what cross-validationschemes were used to estimate model performance. **TabS18.** Summary for feature selection methods in existing studies. **TabS19.** Summary for what neural featureswere used in existing studies. **TabS20.** Summary for what pre-processing methods were used in existing studies. **TabS21.** Trends for the ratio of using open dataset on training ML models. **TabS22.** Results for comparison between SVM and DL classifiers on model performance. **TabS23.** Results for comparison between external validation CVCV) and otherson model performance. **TabS24.** Results for correlation between time and quality scores. **TabS25.** Study quality across psychiatric category.

Additional file 2: Table S1. Evidence table. This table is to summary the metadata and metainformation of all the included studies.

Additional file 3: Table S1-S2. **TabS1.** RNIMP 2022 Checklist. Reporting guideline and checklist for neuroimaging-based machine learning studies for psychiatry. **TabS2.** RNIMP 2022 workflow diagram. A workflow to guide for reporting neuroimaging-based machine learning models for psychiatry.

Authors' contributions

ZC: conceptualization, methodology, software, visualization, and writing—original draft; BH: data curation, methodology, and validation; XL: data curation, methodology, and validation BB: writing—review and editing, and methodology; SE: writing—review and editing, and conceptualization; KM, XG, and YT: validation and data curation; ZX: validation and data curation; XD: validation and data curation; CL: conceptualization, validation, data curation; AL: validation and data curation; JC: conceptualization, validation, and writing—revision and editing; ZF: project administration and funding acquisition; HCP: conceptualization, validation, writing—revision and editing, and supervision. All authors read and approved the final manuscript.

Chen *et al. BMC Medicine*   (2023) 21:241

Page 16 of 29

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the Open Science Framework (OSF) repository, https://osf.io/4zhsp/.

Declarations

Ethics approval and consent to participate

The IRB of the Third Military Medical University (TMMU) exempted full-length censors towards the current study because it is not involved original human or animal data.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Experimental Research Center for Medical and Psychological Science (ERC-MPS), School of Psychology, Third Military Medical University, Chongqing, China. [2]Faculty of Psychology, Southwest University, Chongqing, China. [3]The Center of Psychosomatic Medicine, Sichuan Provincial Center for Mental Health, Sichuan Provincial People's Hospital, Chengdu, China. [4]The Clinical Hospital of Chengdu Brain Science Institute, MOE Key Laboratory for Neuroinformation, University of Electronic Science and Technology of China, Chengdu, China. [5]Institute of Systems Neuroscience, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [6]School of Business and Management, Shanghai International Studies University, Shanghai, China. [7]Department of Radiology, The Third Affiliated Hospital, Sun Yat-Sen University, Guangdong, China. [8]School of Psychology, Clark University, Worcester, MA, USA. [9]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China. [10]Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou, China. [11]Department of Psychiatry, The Fourth Affiliated Hospital, Zhejiang University School of Medicine, Yiwu, Zhejiang, China. [12]School of Psychology, Nanjing Normal University, Nanjing, China.

References

1. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science. 2015;349(6245):255–60.
2. Eyre HA, Singh AB, Reynolds C 3rd. Tech giants enter mental health. World Psychiatry. 2016;15(1):21–2.
3. Walter M, Alizadeh S, Jamalabadi H, Lueken U, Dannlowski U, Walter H, Olbrich S, Colic L, Kambeitz J, Koutsouleris N, et al. Translational machine learning for psychiatric neuroimaging. Prog Neuropsychopharmacol Biol Psychiatry. 2019;91:113–21.
4. Rutherford S. The promise of machine learning for psychiatry. Biol Psychiatry. 2020;88(11):e53–5.
5. Sui J, Jiang R, Bustillo J, Calhoun V. Neuroimaging-based individualized prediction of cognition and behavior for mental disorders and health: methods and promises. Biol Psychiatry. 2020;88(11):818–28.
6. Bzdok D, Varoquaux G, Steyerberg EW. Prediction, not association, paves the road to precision medicine. JAMA Psychiat. 2021;78(2):127–8.
7. Vayena E, Blasimme A. A systemic approach to the oversight of machine learning clinical translation. Am J Bioeth. 2022;22(5):23–5.
8. Poldrack RA, Huckins G, Varoquaux G. Establishment of best practices for evidence for prediction: a review. JAMA Psychiat. 2020;77(5):534–40.
9. Nielsen AN, Barch DM, Petersen SE, Schlaggar BL, Greene DJ. Machine learning with neuroimaging: evaluating its applications in psychiatry. Biol Psychiatry Cogn Neurosci Neuroimaging. 2020;5(8):791–8.
10. 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;145(Pt B):166–79.
11. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. Annu Rev Clin Psychol. 2018;14:91–118.
12. Mihalik A, Ferreira FS, Moutoussis M, Ziegler G, Adams RA, Rosa MJ, Prabhu G, de Oliveira L, Pereira M, Bullmore ET, et al. Multiple holdouts with stability: improving the generalizability of machine learning analyses of brain-behavior relationships. Biol Psychiatry. 2020;87(4):368–76.
13. Meehan AJ, Lewis SJ, Fazel S, Fusar-Poli P, Steyerberg EW, Stahl D, Danese A. Clinical prediction models in psychiatry: a systematic review of two decades of progress and challenges. Mol Psychiatry. 2022;27(6):2700–8.
14. Yuan J, Ran X, Liu K, Yao C, Yao Y, Wu H, Liu Q. Machine learning applications on neuroimaging for diagnosis and prognosis of epilepsy: A review. J Neurosci Methods. 2022;368:109441.
15. Davatzikos C. Machine learning in neuroimaging: Progress and challenges. Neuroimage. 2019;197:652–6.
16. Shrout PE, Rodgers JL. Psychology, science, and knowledge construction: broadening perspectives from the replication crisis. Annu Rev Psychol. 2018;69:487–510.
17. Maxwell SE, Lau MY, Howard GS. Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? Am Psychol. 2015;70(6):487–98.
18. Henrich J, Heine SJ, Norenzayan A. Most people are not WEIRD. Nature. 2010;466(7302):29.
19. Muthukrishna M, Bell AV, Henrich J, Curtin CM, Gedranovich A, McInerney J, Thue B. Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) psychology: measuring and mapping scales of cultural and psychological distance. Psychol Sci. 2020;31(6):678–701.
20. Rad MS, Martingano AJ, Ginges J. Toward a psychology of Homo sapiens: making psychological science more representative of the human population. Proc Natl Acad Sci U S A. 2018;115(45):11401–5.
21. Arnett JJ. The neglected 95%: why American psychology needs to become less American. Am Psychol. 2008;63(7):602–14.
22. Chen Z, Liu X, Yang Q, Wang YJ, Miao K, Gong Z, Yu Y, Leonov A, Liu C, Feng Z, 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.
23. Cearns M, Hahn T, Baune BT. Recommendations and future directions for supervised machine learning in psychiatry. Transl Psychiatry. 2019;9(1):271.
24. Tiwari P, Verma R. The pursuit of generalizability to enable clinical translation of radiomics. Radiol Artif Intell. 2021;3(1):e200227.
25. Wolfers T, Doan NT, Kaufmann T, Alnæs D, Moberget T, Agartz I, Buitelaar JK, Ueland T, Melle I, Franke B, et al. Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. JAMA Psychiat. 2018;75(11):1146–55.
26. Schultebraucks K, Choi KW, Galatzer-Levy IR, Bonanno GA. Discriminating heterogeneous trajectories of resilience and depression after major life stressors using polygenic scores. JAMA Psychiat. 2021;78(7):744–52.
27. Lee HB, Lyketsos CG. Depression in Alzheimer's disease: heterogeneity and related issues. Biol Psychiatry. 2003;54(3):353–62.
28. Arguello PA, Gogos JA. Genetic and cognitive windows into circuit mechanisms of psychiatric disease. Trends Neurosci. 2012;35(1):3–13.
29. Ying X. An overview of overfitting and its solutions. J Phys: Conf Ser. 2019;1168(2):022022.
30. Peng Y, Nagata MH. An empirical overview of nonlinearity and overfitting in machine learning using COVID-19 data. Chaos, Solitons Fractals. 2020;139:110055.
31. Andaur Navarro CL, Damen JAA, Takada T, Nijman SWJ, Dhiman P, Ma J, Collins GS, Bajpai R, Riley RD, Moons KGM, et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. BMJ. 2021;375:n2281.
32. Hosseini M, Powell M, Collins J, Callahan-Flintoft C, Jones W, Bowman H, Wyble B. I tried a bunch of things: the dangers of unexpected overfitting in classification of brain data. Neurosci Biobehav Rev. 2020;119:456–67.

Chen *et al. BMC Medicine*   (2023) 21:241

Page 17 of 29

33. Varoquaux G, Cheplygina V. Machine learning for medical imaging: methodological failures and recommendations for the future. NPJ Digit Med. 2022;5(1):48.

34. Varoquaux G. Cross-validation failure: small sample sizes lead to large error bars. Neuroimage. 2018;180(Pt A):68–77.

35. Chen J, Patil KR, Yeo BTT, Eickhoff SB. Leveraging machine learning for gaining neurobiological and nosological insights in psychiatric research. Biol Psychiatry. 2023;93(1):18–28.

36. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.

37. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, Reitsma JB, Collins GS. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 2014;11(10):e1001744.

38. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350:g7594.

39. Nam CW. World Economic Outlook for 2020 and 2021. CESifo Forum. 2020;21(2):58-9. https://www.proquest.com/openview/2b714d1282 ff098661c0d252c4db128b/1?cbl=43805&pq-origsite=gscholar&paren tSessionId=7a4xwuy%2B60cPGopgOGEQ6SUez3gxXxwiOjjkxULCRuI% 3D.

40. How does the world bank classify countries? https://papers.ssrn.com/ sol3/papers.cfm?abstract_id=2741183.

41. Dagum C. A new approach to the decomposition of the Gini income inequality ratio. Empir Econ. 1997;22:515–31.

42. Hamed KH, Ramachandra Rao A. A modified Mann-Kendall trend test for autocorrelated data. J Hydrol. 1998;204(1):182–96.

43. Battle DE. Diagnostic and Statistical Manual of Mental Disorders (DSM). CoDAS. 2013;25(2):191–2.

44. Huang Y, Wang Y, Wang H, Liu Z, Yu X, Yan J, Yu Y, Kou C, Xu X, Lu J, et al. Prevalence of mental disorders in China: a cross-sectional epidemiological study. Lancet Psychiatry. 2019;6(3):211–24.

45. Ormel J, VonKorff M. Reducing common mental disorder prevalence in populations. JAMA Psychiat. 2021;78(4):359–60.

46. Flint C, Cearns M, Opel N, Redlich R, Mehler DMA, Emden D, Winter NR, Leenings R, Eickhoff SB, Kircher T, et al. Systematic misestimation of machine learning performance in neuroimaging studies of depression. Neuropsychopharmacology. 2021;46(8):1510–7.

47. Woo CW, Chang LJ, Lindquist MA, Wager TD. Building better biomarkers: brain models in translational neuroimaging. Nat Neurosci. 2017;20(3):365–77.

48. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, Logullo P, Beam AL, Peng L, Van Calster B, 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.

49. Poldrack RA, Baker CI, Durnez J, Gorgolewski KJ, Matthews PM, Munafò MR, Nichols TE, Poline JB, Vul E, Yarkoni T. Scanning the horizon: towards transparent and reproducible neuroimaging research. Nat Rev Neurosci. 2017;18(2):115–26.

50. Knoth IS, Lajnef T, Rigoulot S, Lacourse K, Vannasing P, Michaud JL, Jacquemont S, Major P, Jerbi K, Lippé S. Auditory repetition suppression alterations in relation to cognitive functioning in fragile X syndrome: a combined EEG and machine learning approach. J Neurodev Disord. 2018;10(1):4.

51. Pedersen M, Curwood EK, Archer JS, Abbott DF, Jackson GD. Brain regions with abnormal network properties in severe epilepsy of Lennox-Gastaut phenotype: multivariate analysis of task-free fMRI. Epilepsia. 2015;56(11):1767–73.

52. Wang Y, Yuan L, Shi J, Greve A, Ye J, Toga AW, Reiss AL, Thompson PM. Applying tensor-based morphometry to parametric surfaces can improve MRI-based disease diagnosis. Neuroimage. 2013;74:209–30.

53. Hoeft F, Walter E, Lightbody AA, Hazlett HC, Chang C, Piven J, Reiss AL. Neuroanatomical differences in toddler boys with fragile x syndrome and idiopathic autism. Arch Gen Psychiatry. 2011;68(3):295–305.

54. Parisot S, Ktena SI, Ferrante E, Lee M, Guerrero R, Glocker B, Rueckert D. Disease prediction using graph convolutional networks: application to autism spectrum disorder and Alzheimer's disease. Med Image Anal. 2018;48:117–30.

55. Matlis S, Boric K, Chu CJ, Kramer MA. Robust disruptions in electroencephalogram cortical oscillations and large-scale functional networks in autism. BMC Neurol. 2015;15:97.

56. Ingalhalikar M, Parker D, Bloy L, Roberts TP, Verma R. Diffusion based abnormality markers of pathology: toward learned diagnostic prediction of ASD. Neuroimage. 2011;57(3):918–27.

57. Shahamat H, Saniee Abadeh M. Brain MRI analysis using a deep learning based evolutionary approach. Neural Netw. 2020;126:218–34.

58. Bajestani GS, Behrooz M, Khani AG, Nouri-Baygi M, Mollaei A. Diagnosis of autism spectrum disorder based on complex network features. Comput Methods Programs Biomed. 2019;177:277–83.

59. Lanka P, Rangaprakash D, Dretsch MN, Katz JS, Denney TS Jr, Deshpande G. Supervised machine learning for diagnostic classification from large-scale neuroimaging datasets. Brain Imaging Behav. 2020;14(6):2378–416.

60. Zhang L, Wang XH, Li L. Diagnosing autism spectrum disorder using brain entropy: a fast entropy method. Comput Methods Programs Biomed. 2020;190:105240.

61. Rakić M, Cabezas M, Kushibar K, Oliver A, Lladó X. Improving the detection of autism spectrum disorder by combining structural and functional MRI information. NeuroImage Clin. 2020;25:102181.

62. Ahmadlou M, Adeli H, Adeli A. Fractality and a wavelet-chaos-neural network methodology for EEG-based diagnosis of autistic spectrum disorder. J Clin Neurophysiol. 2010;27(5):328–33.

63. Libero LE, DeRamus TP, Lahti AC, Deshpande G, Kana RK. Multimodal neuroimaging based classification of autism spectrum disorder using anatomical, neurochemical, and white matter correlates. Cortex. 2015;66:46–59.

64. Pham TH, Vicnesh J, Wei JKE, Oh SL, Arunkumar N, Abdulhay EW, Ciaccio EJ, Acharya UR. Autism Spectrum Disorder Diagnostic System Using HOS Bispectrum with EEG Signals. Int J Environ Res Public Health. 2020;17(3):971.

65. Graa O, Rekik I. Multi-view learning-based data proliferator for boosting classification using highly imbalanced classes. J Neurosci Methods. 2019;327:108344.

66. Ingalhalikar M, Smith AR, Bloy L, Gur R, Roberts TP, Verma R. Identifying sub-populations via unsupervised cluster analysis on multi-edge similarity graphs. Med Image Comput Comput Assist Interv. 2012;15(Pt 2):254–61.

67. Khosla M, Jamison K, Kuceyeski A, Sabuncu MR. Ensemble learning with 3D convolutional neural networks for functional connectome-based prediction. Neuroimage. 2019;199:651–62.

68. Li H, Parikh NA, He L. A novel transfer learning approach to enhance deep neural network classification of brain functional connectomes. Front Neurosci. 2018;12:491.

69. Sen B, Borle NC, Greiner R, Brown MRG. A general prediction model for the detection of ADHD and Autism using structural and functional MRI. PLoS One. 2018;13(4):e0194856.

70. Xu L, Hua Q, Yu J, Li J. Classification of autism spectrum disorder based on sample entropy of spontaneous functional near infra-red spectroscopy signal. Clin Neurophysiol. 2020;131(6):1365–74.

71. Ma X, Wang XH, Li L. Identifying individuals with autism spectrum disorder based on the principal components of whole-brain phase synchrony. Neurosci Lett. 2021;742:135519.

72. Rakhimberdina Z, Liu X, Murata AT. Population graph-based multimodel ensemble method for diagnosing autism spectrum disorder. Sensors (Basel). 2020;20(21):6001.

73. Tsiaras V, Simos PG, Rezaie R, Sheth BR, Garyfallidis E, Castillo EM, Papanicolaou AC. Extracting biomarkers of autism from MEG resting-state functional connectivity networks. Comput Biol Med. 2011;41(12):1166–77.

74. Wang H, Chen C, Fushing H. Extracting multiscale pattern information of fMRI based functional brain connectivity with application on classification of autism spectrum disorders. PLoS One. 2012;7(10):e45502.

75. Hu J, Cao L, Li T, Liao B, Dong S, Li P. Interpretable learning approaches in resting-state functional connectivity analysis: the case of autism spectrum disorder. Comput Math Methods Med. 2020;2020:1394830.

76. Jung M, Tu Y, Park J, Jorgenson K, Lang C, Song W, Kong J. Surfacebased shared and distinct resting functional connectivity in

Chen *et al. BMC Medicine*   (2023) 21:241

Page 18 of 29

attention-deficit hyperactivity disorder and autism spectrum disorder. Br J Psychiatry. 2019;214(6):339–44.

77. Heinsfeld AS, Franco AR, Craddock RC, Buchweitz A, Meneguzzi F. Identification of autism spectrum disorder using deep learning and the ABIDE dataset. NeuroImage Clinical. 2018;17:16–23.

78. Bhaumik R, Pradhan A, Das S, Bhaumik DK. Predicting autism spectrum disorder using domain-adaptive cross-site evaluation. Neuroinformatics. 2018;16(2):197–205.

79. Wang L, Wee CY, Tang X, Yap PT, Shen D. Multi-task feature selection via supervised canonical graph matching for diagnosis of autism spectrum disorder. Brain Imaging Behav. 2016;10(1):33–40.

80. Ecker C, Rocha-Rego V, Johnston P, Mourao-Miranda J, Marquand A, Daly EM, Brammer MJ, Murphy C, Murphy DG. Investigating the predictive value of whole-brain structural MR scans in autism: a pattern classification approach. Neuroimage. 2010;49(1):44–56.

81. Payabvash S, Palacios EM, Owen JP, Wang MB, Tavassoli T, Gerdes M, Brandes-Aitken A, Cuneo D, Marco EJ, Mukherjee P. White matter connectome edge density in children with autism spectrum disorders: potential imaging biomarkers using machine-learning models. Brain Connect. 2019;9(2):209–20.

82. Price T, Wee CY, Gao W, Shen D. Multiple-network classification of childhood autism using functional connectivity dynamics. Med Image Comput Comput Assist Interv. 2014;17(Pt 3):177–84.

83. Haweel R, Shalaby A, Mahmoud A, Seada N, Ghoniemy S, Ghazal M, Casanova MF, Barnes GN, El-Baz A. A robust DWT-CNN-based CAD system for early diagnosis of autism using task-based fMRI. Med Phys. 2021;48(5):2315–26.

84. Chen CP, Keown CL, Jahedi A, Nair A, Pflieger ME, Bailey BA, Müller RA. Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual regions in autism. NeuroImage Clin. 2015;8:238–45.

85. Anderson JS, Nielsen JA, Froehlich AL, DuBray MB, Druzgal TJ, Cariello AN, Cooperrider JR, Zielinski BA, Ravichandran C, Fletcher PT, et al. Functional connectivity magnetic resonance imaging classification of autism. Brain. 2011;134(Pt 12):3742–54.

86. Uddin LQ, Supekar K, Lynch CJ, Khouzam A, Phillips J, Feinstein C, Ryali S, Menon V. Salience network-based classification and prediction of symptom severity in children with autism. JAMA Psychiat. 2013;70(8):869–79.

87. Nielsen JA, Zielinski BA, Fletcher PT, Alexander AL, Lange N, Bigler ED, Lainhart JE, Anderson JS. Multisite functional connectivity MRI classification of autism: ABIDE results. Front Hum Neurosci. 2013;7:599.

88. Jahedi A, Nasamran CA, Faires B, Fan J, Müller RA. Distributed intrinsic functional connectivity patterns predict diagnostic status in Large Autism Cohort. Brain Connect. 2017;7(8):515–25.

89. Retico A, Giuliano A, Tancredi R, Cosenza A, Apicella F, Narzisi A, Biagi L, Tosetti M, Muratori F, Calderoni S. The effect of gender on the neuroanatomy of children with autism spectrum disorders: a support vector machine case-control study. Mol Autism. 2016;7:5.

90. Calderoni S, Retico A, Biagi L, Tancredi R, Muratori F, Tosetti M. Female children with autism spectrum disorder: an insight from mass-univariate and pattern classification analyses. Neuroimage. 2012;59(2):1013–22.

91. Yamagata B, Itahashi T, Fujino J, Ohta H, Nakamura M, Kato N, Mimura M, Hashimoto RI, Aoki Y. Machine learning approach to identify a resting-state functional connectivity pattern serving as an endophenotype of autism spectrum disorder. Brain Imaging Behav. 2019;13(6):1689–98.

92. Gori I, Giuliano A, Muratori F, Saviozzi I, Oliva P, Tancredi R, Cosenza A, Tosetti M, Calderoni S, Retico A. Gray matter alterations in young children with autism spectrum disorders: comparing morphometry at the voxel and regional level. J Neuroimaging. 2015;25(6):866–74.

93. Leming M, Górriz JM, Suckling J. Ensemble deep learning on large, mixed-site fMRI datasets in autism and other tasks. Int J Neural Syst. 2020;30(7):2050012.

94. Shen MD, Nordahl CW, Li DD, Lee A, Angkustsiri K, Emerson RW, Rogers SJ, Ozonoff S, Amaral DG. Extra-axial cerebrospinal fluid in high-risk and normal-risk children with autism aged 2–4 years: a case-control study. The lancet Psychiatry. 2018;5(11):895–904.

95. Grossi E, Olivieri C, Buscema M. Diagnosis of autism through EEG processed by advanced computational algorithms: a pilot study. Comput Methods Programs Biomed. 2017;142:73–9.

96. Gupta S, Rajapakse JC, Welsch RE. Ambivert degree identifies crucial brain functional hubs and improves detection of Alzheimer's Disease and Autism Spectrum Disorder. NeuroImage Clinical. 2020;25:102186.

97. Ghiassian S, Greiner R, Jin P, Brown MR. Using functional or structural magnetic resonance images and personal characteristic data to identify ADHD and Autism. PLoS One. 2016;11(12):e0166934.

98. Zu C, Gao Y, Munsell B, Kim M, Peng Z, Cohen JR, Zhang D, Wu G. Identifying disease-related subnetwork connectome biomarkers by sparse hypergraph learning. Brain Imaging Behav. 2019;13(4):879–92.

99. Katuwal GJ, Baum SA, Cahill ND, Michael AM. Divide and Conquer: Sub-Grouping of ASD Improves ASD Detection Based on Brain Morphometry. PLoS One. 2016;11(4):e0153331.

100. Li Q, Becker B, Jiang X, Zhao Z, Zhang Q, Yao S, Kendrick KM. Decreased interhemispheric functional connectivity rather than corpus callosum volume as a potential biomarker for autism spectrum disorder. Cortex. 2019;119:258–66.

101. Dekhil O, Hajjdiab H, Shalaby A, Ali MT, Ayinde B, Switala A, Elshamekh A, Ghazal M, Keynton R, Barnes G, et al. Using resting state functional MRI to build a personalized autism diagnosis system. PLoS One. 2018;13(10):e0206351.

102. Yamagata B, Itahashi T, Fujino J, Ohta H, Takashio O, Nakamura M, Kato N, Mimura M, Hashimoto RI, Aoki YY. Cortical surface architecture endophenotype and correlates of clinical diagnosis of autism spectrum disorder. Psychiatry Clin Neurosci. 2019;73(7):409–15.

103. Iidaka T. Resting state functional magnetic resonance imaging and neural network classified autism and control. Cortex. 2015;63:55–67.

104. Jamal W, Das S, Oprescu IA, Maharatna K, Apicella F, Sicca F. Classification of autism spectrum disorder using supervised learning of brain connectivity measures extracted from synchrostates. J Neural Eng. 2014;11(4):046019.

105. Zhang F, Savadjiev P, Cai W, Song Y, Rathi Y, Tunç B, Parker D, Kapur T, Schultz RT, Makris N, et al. Whole brain white matter connectivity analysis using machine learning: an application to autism. Neuroimage. 2018;172:826–37.

106. Sujit SJ, Coronado I, Kamali A, Narayana PA, Gabr RE. Automated image quality evaluation of structural brain MRI using an ensemble of deep learning networks. J Magn Reson Imaging. 2019;50(4):1260–7.

107. Huang H, Liu X, Jin Y, Lee SW, Wee CY, Shen D. Enhancing the representation of functional connectivity networks by fusing multi-view information for autism spectrum disorder diagnosis. Hum Brain Mapp. 2019;40(3):833–54.

108. Eill A, Jahedi A, Gao Y, Kohli JS, Fong CH, Solders S, Carper RA, Valafar F, Bailey BA, Müller RA. Functional connectivities are more informative than anatomical variables in diagnostic classification of autism. Brain Connect. 2019;9(8):604–12.

109. Xiao X, Fang H, Wu J, Xiao C, Xiao T, Qian L, Liang F, Xiao Z, Chu KK, Ke X. Diagnostic model generated by MRI-derived brain features in toddlers with autism spectrum disorder. Autism Res. 2017;10(4):620–30.

110. Kam TE, Suk HI, Lee SW. Multiple functional networks modeling for autism spectrum disorder diagnosis. Hum Brain Mapp. 2017;38(11):5804–21.

111. Ktena SI, Parisot S, Ferrante E, Rajchl M, Lee M, Glocker B, Rueckert D. Metric learning with spectral graph convolutions on brain connectivity networks. Neuroimage. 2018;169:431–42.

112. Aghdam MA, Sharifi A, Pedram MM. Diagnosis of autism spectrum disorders in young children based on resting-state functional magnetic resonance imaging data using convolutional neural networks. J Digit Imaging. 2019;32(6):899–918.

113. Sadeghi M, Khosrowabadi R, Bakouie F, Mahdavi H, Eslahchi C, Pouretemad H. Screening of autism based on task-free fMRI using graph theoretical approach. Psychiatry Res Neuroimaging. 2017;263:48–56.

114. Chaddad A, Desrosiers C, Hassan L, Tanougast C. Hippocampus and amygdala radiomic biomarkers for the study of autism spectrum disorder. BMC Neurosci. 2017;18(1):52.

115. Djemal R, AlSharabi K, Ibrahim S, Alsuwailem A. EEG-based computer aided diagnosis of autism spectrum disorder using wavelet, entropy, and ANN. Biomed Res Int. 2017;2017:9816591.

Chen *et al. BMC Medicine*   (2023) 21:241

Page 19 of 29

116. Yahata N, Morimoto J, Hashimoto R, Lisi G, Shibata K, Kawakubo Y, Kuwabara H, Kuroda M, Yamada T, Megumi F, et al. A small number of abnormal brain connections predicts adult autism spectrum disorder. Nat Commun. 2016;7:11254.

117. Heunis T, Aldrich C, Peters JM, Jeste SS, Sahin M, Scheffer C, de Vries PJ. Recurrence quantification analysis of resting state EEG signals in autism spectrum disorder - a systematic methodological exploration of technical and demographic confounders in the search for biomarkers. BMC Med. 2018;16(1):101.

118. Chen H, Duan X, Liu F, Lu F, Ma X, Zhang Y, Uddin LQ, Chen H. Multivariate classification of autism spectrum disorder using frequency-specific resting-state functional connectivity–A multi-center study. Prog Neuropsychopharmacol Biol Psychiatry. 2016;64:1–9.

119. Just MA, Cherkassky VL, Buchweitz A, Keller TA, Mitchell TM. Identifying autism from neural representations of social interactions: neurocognitive markers of autism. PLoS One. 2014;9(12):e113879.

120. Akhavan Aghdam M, Sharifi A, Pedram MM. Combination of rs-fMRI and sMRI data to discriminate autism spectrum disorders in young children using deep belief network. J Digit Imaging. 2018;31(6):895–903.

121. Alturki FA, AlSharabi K, Abdurraqeeb AM, Aljalal M. EEG signal analysis for diagnosing neurological disorders using discrete wavelet transform and intelligent techniques. Sensors (Basel). 2020;20(9):2505.

122. Spiegel A, Mentch J, Haskins AJ, Robertson CE. Slower binocular rivalry in the autistic brain. Curr Biol. 2019;29(17):2948-2953.e2943.

123. Conti E, Retico A, Palumbo L, Spera G, Bosco P, Biagi L, Fiori S, Tosetti M, Cipriani P, Cioni G, et al. Autism spectrum disorder and childhood apraxia of speech: early language-related hallmarks across structural MRI study. J Pers Med. 2020;10(4):275.

124. Bi XA, Liu Y, Jiang Q, Shu Q, Sun Q, Dai J. The diagnosis of autism spectrum disorder based on the random neural network cluster. Front Hum Neurosci. 2018;12:257.

125. Pollonini L, Patidar U, Situ N, Rezaie R, Papanicolaou AC, Zouridakis G. Functional connectivity networks in the autistic and healthy brain assessed using Granger causality. Annu Int Conf IEEE Eng Med Biol Soc. 2010;2010:1730–3.

126. Eldridge J, Lane AE, Belkin M, Dennis S. Robust features for the automatic identification of autism spectrum disorder in children. J Neurodev Disord. 2014;6(1):12.

127. Khan NA, Waheeb SA, Riaz A, Shang X. A three-stage teacher, student neural networks and sequential feed forward selection-based feature selection approach for the classification of autism spectrum disorder. Brain Sci. 2020;10(10):754.

128. Sherkatghanad Z, Akhondzadeh M, Salari S, Zomorodi-Moghadam M, Abdar M, Acharya UR, Khosrowabadi R, Salari V. Automated detection of autism spectrum disorder using a convolutional neural network. Front Neurosci. 2019;13:1325.

129. Gao J, Chen M, Li Y, Gao Y, Li Y, Cai S, Wang J. Multisite autism spectrum disorder classification using convolutional neural network classifier and individual morphological brain networks. Front Neurosci. 2020;14:629630.

130. Liu Y, Xu L, Li J, Yu J, Yu X. Attentional connectivity-based prediction of autism using heterogeneous rs-fMRI data from CC200 Atlas. Exp Neurobiol. 2020;29(1):27–37.

131. Yang M, Cao M, Chen Y, Chen Y, Fan G, Li C, Wang J, Liu T. Large-scale brain functional network integration for discrimination of autism using a 3-D Deep Learning Model. Front Hum Neurosci. 2021;15:687288.

132. Huang ZA, Zhu Z, Yau CH, Tan KC. Identifying autism spectrum disorder from resting-state fMRI using deep belief network. IEEE Trans Neural Netw Learn Syst. 2021;32(7):2847–61.

133. Zhao J, Song J, Li X, Kang J. A study on EEG feature extraction and classification in autistic children based on singular spectrum analysis method. Brain Behav. 2020;10(12):e01721.

134. Almuqhim F, Saeed F. ASD-SAENet: A Sparse Autoencoder, and Deep-Neural Network Model for Detecting Autism Spectrum Disorder (ASD) Using fMRI Data. Front Comput Neurosci. 2021;15:654315.

135. Xu L, Sun Z, Xie J, Yu J, Li J, Wang J. Identification of autism spectrum disorder based on short-term spontaneous hemodynamic fluctuations using deep learning in a multi-layer neural network. Clin Neurophysiol. 2021;132(2):457–68.

136. Lu J, Kishida K, De Asis CJ, Lohrenz T, Deering DT, Beauchamp M, Montague PR. Single stimulus fMRI produces a neural individual difference measure for Autism Spectrum Disorder. Clin Psychol Sci. 2015;3(3):422–32.

137. Ahmed MR, Zhang Y, Liu Y, Liao H. Single volume image generator and deep learning-based ASD classification. IEEE J Biomed Health Inform. 2020;24(11):3044–54.

138. Xu L, Geng X, He X, Li J, Yu J. Prediction in autism by deep learning short-time spontaneous hemodynamic fluctuations. Front Neurosci. 2019;13:1120.

139. Wee CY, Wang L, Shi F, Yap PT, Shen D. Diagnosis of autism spectrum disorders using regional and interregional morphological features. Hum Brain Mapp. 2014;35(7):3414–30.

140. Sewani H, Kashef R. An autoencoder-based deep learning classifier for efficient diagnosis of autism. Children (Basel, Switzerland). 2020;7(10):182.

141. Shi C, Xin X, Zhang J. Domain adaptation using a three-way decision improves the identification of autism patients from multisite fMRI data. Brain Sci. 2021;11(5):603.

142. Kazeminejad A, Sotero RC. The importance of anti-correlations in graph theory based classification of autism spectrum disorder. Front Neurosci. 2020;14:676.

143. Yin W, Mostafa S, Wu FX. Diagnosis of autism spectrum disorder based on functional brain networks with deep learning. J Comput Biol. 2021;28(2):146–65.

144. Jiao Y, Chen R, Ke X, Chu K, Lu Z, Herskovits EH. Predictive models of autism spectrum disorder based on brain regional cortical thickness. Neuroimage. 2010;50(2):589–99.

145. Murdaugh DL, Shinkareva SV, Deshpande HR, Wang J, Pennick MR, Kana RK. Differential deactivation during mentalizing and classification of autism based on default mode network connectivity. PLoS One. 2012;7(11):e50064.

146. Song Y, Epalle TM, Lu H. Characterizing and predicting autism spectrum disorder by performing resting-state functional network community pattern analysis. Front Hum Neurosci. 2019;13:203.

147. Irimia A, Lei X, Torgerson CM, Jacokes ZJ, Abe S, Van Horn JD. Support vector machines, multidimensional scaling and magnetic resonance imaging reveal structural brain abnormalities associated with the interaction between autism spectrum disorder and sex. Front Comput Neurosci. 2018;12:93.

148. Eslami T, Mirjalili V, Fong A, Laird AR, Saeed F. ASD-DiagNet: a hybrid learning approach for detection of autism spectrum disorder using fMRI data. Front Neuroinform. 2019;13:70.

149. Sarovic D, Hadjikhani N, Schneiderman J, Lundström S, Gillberg C. Autism classified by magnetic resonance imaging: a pilot study of a potential diagnostic tool. Int J Methods Psychiatr Res. 2020;29(4):1–18.

150. Kazeminejad A, Sotero RC. Topological properties of resting-state fMRI functional networks improve machine learning-based autism classification. Front Neurosci. 2018;12:1018.

151. Guo X, Dominick KC, Minai AA, Li H, Erickson CA, Lu LJ. Diagnosing autism spectrum disorder from brain resting-state functional connectivity patterns using a deep neural network with a novel feature selection method. Front Neurosci. 2017;11:460.

152. Thomas RM, Gallo S, Cerliani L, Zhutovsky P, El-Gazzar A, van Wingen G. Classifying autism spectrum disorder using the temporal statistics of resting-state functional MRI data with 3D convolutional neural networks. Front Psych. 2020;11:440.

153. Chen H, Chen W, Song Y, Sun L, Li X. EEG characteristics of children with attention-deficit/hyperactivity disorder. Neuroscience. 2019;406:444–56.

154. Guo X, Yao D, Cao Q, Liu L, Zhao Q, Li H, Huang F, Wang Y, Qian Q, Wang Y, et al. Shared and distinct resting functional connectivity in children and adults with attention-deficit/hyperactivity disorder. Transl Psychiatry. 2020;10(1):65.

155. Müller A, Vetsch S, Pershin I, Candrian G, Baschera GM, Kropotov JD, Kasper J, Rehim HA, Eich D. EEG/ERP-based biomarker/neuroalgorithms in adults with ADHD: Development, reliability, and application in clinical practice. World J Biol Psychiatry. 2020;21(3):172–82.

156. Gao MS, Tsai FS, Lee CC. Learning a phenotypic-attribute attentional brain connectivity embedding for ADHD Classification using rs-fMRI. Annu Int Conf IEEE Eng Med Biol Soc. 2020;2020:5472–5.

Chen *et al. BMC Medicine*   (2023) 21:241

Page 20 of 29

157. Chen Y, Tang Y, Wang C, Liu X, Zhao L, Wang Z. ADHD classification by dual subspace learning using resting-state functional connectivity. Artif Intell Med. 2020;103:101786.

158. Muthuraman M, Moliadze V, Boecher L, Siemann J, Freitag CM, Groppa S, Siniatchkin M. Multimodal alterations of directed connectivity profiles in patients with attention-deficit/hyperactivity disorders. Sci Rep. 2019;9(1):20028.

159. McNorgan C, Judson C, Handzlik D, Holden JG. Linking ADHD and behavioral assessment through identification of shared diagnostic task-based functional connections. Front Physiol. 2020;11:583005.

160. Vahid A, Bluschke A, Roessner V, Stober S, Beste C. Deep learning based on event-related EEG differentiates children with ADHD from Healthy Controls. J Clin Med. 2019;8(7):1055.

161. Riaz A, Asad M, Alonso E, Slabaugh G. DeepFMRI: End-to-end deep learning for functional connectivity and classification of ADHD using fMRI. J Neurosci Methods. 2020;335:108506.

162. Rostami M, Farashi S, Khosrowabadi R, Pouretemad H. Discrimination of ADHD subtypes using decision tree on behavioral, neuropsychological, and neural markers. Basic Clin Neurosci. 2020;11(3):359–67.

163. Kiiski H, Rueda-Delgado LM, Bennett M, Knight R, Rai L, Roddy D, Grogan K, Bramham J, Kelly C, Whelan R. Functional EEG connectivity is a neuromarker for adult attention deficit hyperactivity disorder symptoms. Clin Neurophysiol. 2020;131(1):330–42.

164. Tang Y, Wang C, Chen Y, Sun N, Jiang A, Wang Z. Identifying ADHD individuals from resting-state functional connectivity using subspace clustering and binary hypothesis testing. J Atten Disord. 2021;25(5):736–48.

165. Sun Y, Zhao L, Lan Z, Jia XZ, Xue SW. Differentiating boys with ADHD from those with typical development based on whole-brain functional connections using a machine learning approach. Neuropsychiatr Dis Treat. 2020;16:691–702.

166. Sidhu G. Locally linear embedding and fMRI feature selection in psychiatric classification. IEEE J Transl Eng Health Med. 2019;7:2200211.

167. Riaz A, Asad M, Alonso E, Slabaugh G. Fusion of fMRI and non-imaging data for ADHD classification. Comput Med Imaging Graph. 2018;65:115–28.

168. Kaur S, Singh S, Arun P, Kaur D, Bajaj M. Phase Space Reconstruction of EEG Signals for Classification of ADHD and Control Adults. Clin EEG Neurosci. 2020;51(2):102–13.

169. Chen M, Li H, Wang J, Dillman JR, Parikh NA, He L. A multichannel deep neural network model analyzing multiscale functional brain connectome data for attention deficit hyperactivity disorder detection. Radiol Artif Intell. 2019;2(1):e190012.

170. Sutoko S, Monden Y, Tokuda T, Ikeda T, Nagashima M, Funane T, Sato H, Kiguchi M, Maki A, Yamagata T, et al. Exploring attentive task-based connectivity for screening attention deficit/hyperactivity disorder children: a functional near-infrared spectroscopy study. Neurophotonics. 2019;6(4):045013.

171. Luo Y, Alvarez TL, Halperin JM, Li X. Multimodal neuroimaging-based prediction of adult outcomes in childhood-onset ADHD using ensemble learning techniques. NeuroImage Clin. 2020;26:102238.

172. Wang XH, Jiao Y, Li L. Diagnostic model for attention-deficit hyperactivity disorder based on interregional morphological connectivity. Neurosci Lett. 2018;685:30–4.

173. Peng X, Lin P, Zhang T, Wang J. Extreme learning machine-based classification of ADHD using brain structural MRI data. PLoS One. 2013;8(11):e79476.

174. Yasumura A, Omori M, Fukuda A, Takahashi J, Yasumura Y, Nakagawa E, Koike T, Yamashita Y, Miyajima T, Koeda T, et al. Applied machine learning method to predict children with ADHD using prefrontal cortex activity: a multicenter study in Japan. J Atten Disord. 2020;24(14):2012–20.

175. Qureshi MNI, Oh J, Min B, Jo HJ, Lee B. Multi-modal, multi-measure, and multi-class discrimination of ADHD with hierarchical feature extraction and extreme learning machine using structural and functional brain MRI. Front Hum Neurosci. 2017;11:157.

176. Biederman J, Hammerness P, Sadeh B, Peremen Z, Amit A, Or-Ly H, Stern Y, Reches A, Geva A, Faraone SV. Diagnostic utility of brain activity flow patterns analysis in attention deficit hyperactivity disorder. Psychol Med. 2017;47(7):1259–70.

177. Gehricke JG, Kruggel F, Thampipop T, Alejo SD, Tatos E, Fallon J, Muftuler LT. The brain anatomy of attention-deficit/hyperactivity disorder in young adults - a magnetic resonance imaging study. PLoS One. 2017;12(4):e0175433.

178. Sato JR, Hoexter MQ, Fujita A, Rohde LA. Evaluation of pattern recognition and feature extraction methods in ADHD prediction. Front Syst Neurosci. 2012;6:68.

179. Iannaccone R, Hauser TU, Ball J, Brandeis D, Walitza S, Brem S. Classifying adolescent attention-deficit/hyperactivity disorder (ADHD) based on functional and structural imaging. Eur Child Adolesc Psychiatry. 2015;24(10):1279–89.

180. Gu Y, Miao S, Han J, Liang Z, Ouyang G, Yang J, Li X. Identifying ADHD children using hemodynamic responses during a working memory task measured by functional near-infrared spectroscopy. J Neural Eng. 2018;15(3):035005.

181. Du J, Wang L, Jie B, Zhang D. Network-based classification of ADHD patients using discriminative subnetwork selection and graph kernel PCA. Comput Med Imaging Graph. 2016;52:82–8.

182. Wang XH, Jiao Y, Li L. Identifying individuals with attention deficit hyperactivity disorder based on temporal variability of dynamic functional connectivity. Sci Rep. 2018;8(1):11789.

183. Hart H, Chantiluke K, Cubillo AI, Smith AB, Simmons A, Brammer MJ, Marquand AF, Rubia K. Pattern classification of response inhibition in ADHD: toward the development of neurobiological markers for ADHD. Hum Brain Mapp. 2014;35(7):3083–94.

184. Dai D, Wang J, Hua J, He H. Classification of ADHD children through multimodal magnetic resonance imaging. Front Syst Neurosci. 2012;6:63.

185. Liu S, Zhao L, Wang X, Xin Q, Zhao J, Guttery DS, Zhang YD. Deep spatio-temporal representation and ensemble classification for attention deficit/hyperactivity disorder. IEEE Trans Neural Syst Rehabil Eng. 2021;29:1–10.

186. Wang X, Jiao Y, Tang T, Wang H, Lu Z. Altered regional homogeneity patterns in adults with attention-deficit hyperactivity disorder. Eur J Radiol. 2013;82(9):1552–7.

187. Sidhu GS, Asgarian N, Greiner R, Brown MR. Kernel Principal Component Analysis for dimensionality reduction in fMRI-based diagnosis of ADHD. Front Syst Neurosci. 2012;6:74.

188. Liechti MD, Valko L, Müller UC, Döhnert M, Drechsler R, Steinhausen HC, Brandeis D. Diagnostic value of resting electroencephalogram in attention-deficit/hyperactivity disorder across the lifespan. Brain Topogr. 2013;26(1):135–51.

189. Zhu CZ, Zang YF, Cao QJ, Yan CG, He Y, Jiang TZ, Sui MQ, Wang YF. Fisher discriminative analysis of resting-state brain function for attention-deficit/hyperactivity disorder. Neuroimage. 2008;40(1):110–20.

190. Olivetti E, Greiner S, Avesani P. ADHD diagnosis from multiple data sources with batch effects. Front Syst Neurosci. 2012;6:70.

191. Johnston BA, Mwangi B, Matthews K, Coghill D, Konrad K, Steele JD. Brainstem abnormalities in attention deficit hyperactivity disorder support high accuracy individual diagnostic classification. Hum Brain Mapp. 2014;35(10):5179–89.

192. Aradhya AMS, Subbaraju V, Sundaram S, Sundararajan N. Regularized Spatial Filtering Method (R-SFM) for detection of Attention Deficit Hyperactivity Disorder (ADHD) from resting-state functional Magnetic Resonance Imaging (rs-fMRI). Annu Int Conf IEEE Eng Med Biol Soc. 2018;2018:5541–4.

193. Chang CW, Ho CC, Chen JH. ADHD classification by a texture analysis of anatomical brain MRI data. Front Syst Neurosci. 2012;6:66.

194. Smith JL, Johnstone SJ, Barry RJ. Aiding diagnosis of attention-deficit/hyperactivity disorder and its subtypes: discriminant function analysis of event-related potential data. J Child Psychol Psychiatry. 2003;44(7):1067–75.

195. dos Santos SA, Biazoli Junior CE, Comfort WE, Rohde LA, Sato JR. Abnormal functional resting-state networks in ADHD: graph theory and pattern recognition analysis of fMRI data. Biomed Res Int. 2014;2014:380531.

196. Sun H, Chen Y, Huang Q, Lui S, Huang X, Shi Y, Xu X, Sweeney JA, Gong Q. Psychoradiologic utility of MR imaging for diagnosis of attention deficit hyperactivity disorder: a radiomics analysis. Radiology. 2018;287(2):620–30.

197. Mueller A, Candrian G, Kropotov JD, Ponomarev VA, Baschera GM. Classification of ADHD patients on the basis of independent ERP

Chen *et al. BMC Medicine*   (2023) 21:241

Page 21 of 29

components using a machine learning system. Nonlinear Biomed Phys. 2010;4 Suppl 1(Suppl 1):S1.

198. Cheng W, Ji X, Zhang J, Feng J. Individual classification of ADHD patients by integrating multiscale neuroimaging markers and advanced pattern recognition techniques. Front Syst Neurosci. 2012;6:58.

199. Ahmadlou M, Adeli H. Wavelet-synchronization methodology: a new approach for EEG-based diagnosis of ADHD. Clin EEG Neurosci. 2010;41(1):1–10.

200. Abibullaev B, An J. Decision support algorithm for diagnosis of ADHD using electroencephalograms. J Med Syst. 2012;36(4):2675–88.

201. Colby JB, Rudie JD, Brown JA, Douglas PK, Cohen MS, Shehzad Z. Insights into multimodal imaging classification of ADHD. Front Syst Neurosci. 2012;6:59.

202. Mueller A, Candrian G, Grane VA, Kropotov JD, Ponomarev VA, Baschera GM. Discriminating between ADHD adults and controls using independent ERP components and a support vector machine: a validation study. Nonlinear Biomed Phys. 2011;5:5.

203. Yu D. Additional brain functional network in adults with attention-deficit/hyperactivity disorder: a phase synchrony analysis. PLoS One. 2013;8(1):e54516.

204. Poil SS, Bollmann S, Ghisleni C, O'Gorman RL, Klaver P, Ball J, Eich-Höchli D, Brandeis D, Michels L. Age dependent electroencephalographic changes in attention-deficit/hyperactivity disorder (ADHD). Clin Neurophysiol. 2014;125(8):1626–38.

205. Qureshi MN, Min B, Jo HJ, Lee B. Multiclass classification for the differential diagnosis on the ADHD subtypes using recursive feature elimination and hierarchical extreme learning machine: structural MRI study. PLoS One. 2016;11(8):e0160697.

206. Hart H, Marquand AF, Smith A, Cubillo A, Simmons A, Brammer M, Rubia K. Predictive neurofunctional markers of attention-deficit/hyperactivity disorder based on pattern classification of temporal processing. J Am Acad Child Adolesc Psychiatry. 2014;53(5):569-578.e561.

207. Qureshi MN, Boreom L. Classification of ADHD subgroup with recursive feature elimination for structural brain MRI. Annu Int Conf IEEE Eng Med Biol Soc. 2016;2016:5929–32.

208. Deshpande G, Wang P, Rangaprakash D, Wilamowski B. Fully connected cascade artificial neural network architecture for attention deficit hyperactivity disorder classification from functional magnetic resonance imaging data. IEEE Trans Cybern. 2015;45(12):2668–79.

209. Hammer R, Cooke GE, Stein MA, Booth JR. Functional neuroimaging of visuospatial working memory tasks enables accurate detection of attention deficit and hyperactivity disorder. NeuroImage Clin. 2015;9:244–52.

210. Moghaddari M, Lighvan MZ, Danishvar S. Diagnose ADHD disorder in children using convolutional neural network based on continuous mental task EEG. Comput Methods Programs Biomed. 2020;197:105738.

211. Pereda E, García-Torres M, Melián-Batista B, Mañas S, Méndez L, González JJ. The blessing of dimensionality: feature selection outperforms functional connectivity-based feature transformation to classify ADHD subjects from EEG patterns of phase synchronisation. PLoS One. 2018;13(8):e0201660.

212. Lim L, Marquand A, Cubillo AA, Smith AB, Chantiluke K, Simmons A, Mehta M, Rubia K. Disorder-specific predictive classification of adolescents with attention deficit hyperactivity disorder (ADHD) relative to autism using structural magnetic resonance imaging. PLoS One. 2013;8(5):e63660.

213. Öztoprak H, Toycan M, Alp YK, Arıkan O, Doğutepe E, Karakaş S. Machine-based classification of ADHD and nonADHD participants using time/frequency features of event-related neuroelectric activity. Clin Neurophysiol. 2017;128(12):2400–10.

214. Tor HT, Ooi CP, Lim-Ashworth NS, Wei JKE, Jahmunah V, Oh SL, Acharya UR, Fung DSS. Automated detection of conduct disorder and attention deficit hyperactivity disorder using decomposition and nonlinear techniques with EEG signals. Comput Methods Programs Biomed. 2021;200:105941.

215. Tosun M. Effects of spectral features of EEG signals recorded with different channels and recording statuses on ADHD classification with deep learning. Phys Eng Sci Med. 2021;44(3):693–702.

216. Johnstone SJ, Parrish L, Jiang H, Zhang DW, Williams V, Li S. Aiding diagnosis of childhood attention-deficit/hyperactivity disorder of the

inattentive presentation: Discriminant function analysis of multi-domain measures including EEG. Biol Psychol. 2021;161:108080.

217. Dey S, Rao AR, Shah M. Exploiting the brain's network structure in identifying ADHD subjects. Front Syst Neurosci. 2012;6:75.

218. Yoo JH, Kim JI, Kim BN, Jeong B. Exploring characteristic features of attention-deficit/hyperactivity disorder: findings from multi-modal MRI and candidate genetic data. Brain Imaging Behav. 2020;14(6):2132–47.

219. Tenev A, Markovska-Simoska S, Kocarev L, Pop-Jordanov J, Müller A, Candrian G. Machine learning approach for classification of ADHD adults. Int J Psychophysiol. 2014;93(1):162–6.

220. Rezaeezadeh M, Shamekhi S, Shamsi M. Attention deficit hyperactivity disorder diagnosis using non-linear univariate and multivariate EEG measurements: a preliminary study. Phys Eng Sci Med. 2020;43(2):577–92.

221. Crippa A, Salvatore C, Molteni E, Mauri M, Salandi A, Trabattoni S, Agostoni C, Molteni M, Nobile M, Castiglioni I. The utility of a computerized algorithm based on a multi-domain profile of measures for the diagnosis of attention deficit/hyperactivity disorder. Front Psych. 2017;8:189.

222. Ishii-Takahashi A, Takizawa R, Nishimura Y, Kawakubo Y, Kuwabara H, Matsubayashi J, Hamada K, Okuhata S, Yahata N, Igarashi T, et al. Prefrontal activation during inhibitory control measured by near-infrared spectroscopy for differentiating between autism spectrum disorders and attention deficit hyperactivity disorder in adults. NeuroImage Clin. 2014;4:53–63.

223. Zhang T, Li C, Li P, Peng Y, Kang X, Jiang C, Li F, Zhu X, Yao D, Biswal B, et al. Separated Channel Attention Convolutional Neural Network (SC-CNN-Attention) to Identify ADHD in Multi-Site Rs-fMRI Dataset. Entropy (Basel, Switzerland). 2020;22(8):893.

224. Abramov DM, Lazarev VV, Gomes Junior SC, Mourao-Junior CA, Castro-Pontes M, Cunha CQ, deAzevedo LC, Vigneau E. Estimating biological accuracy of DSM for attention deficit/hyperactivity disorder based on multivariate analysis for small samples. PeerJ. 2019;7:e7074.

225. Dey S, Rao AR, Shah M. Attributed graph distance measure for automatic detection of attention deficit hyperactive disordered subjects. Front Neural Circuits. 2014;8:64.

226. Helgadóttir H, Gudmundsson Ó, Baldursson G, Magnússon P, Blin N, Brynjólfsdóttir B, Emilsdóttir Á, Gudmundsdóttir GB, Lorange M, Newman PK, et al. Electroencephalography as a clinical tool for diagnosing and monitoring attention deficit hyperactivity disorder: a cross-sectional study. BMJ Open. 2015;5(1):e005500.

227. Chow JC, Ouyang CS, Tsai CL, Chiang CT, Yang RC, Wu RC, Wu HC, Lin LC. Entropy-based quantitative electroencephalogram analysis for diagnosing attention-deficit hyperactivity disorder in girls. Clin EEG Neurosci. 2019;50(3):172–9.

228. Chen H, Song Y, Li X. Use of deep learning to detect personalized spatial-frequency abnormalities in EEGs of children with ADHD. J Neural Eng. 2019;16(6):066046.

229. Jahanshahloo HR, Shamsi M, Ghasemi E, Kouhi A. Automated and ERP-based diagnosis of attention-deficit hyperactivity disorder in children. J Med Signals Sens. 2017;7(1):26–32.

230. Wolfers T, van Rooij D, Oosterlaan J, Heslenfeld D, Hartman CA, Hoekstra PJ, Beckmann CF, Franke B, Buitelaar JK, Marquand AF. Quantifying patterns of brain activity: distinguishing unaffected siblings from participants with ADHD and healthy individuals. NeuroImage Clin. 2016;12:227–33.

231. Shao L, You Y, Du H, Fu D. Classification of ADHD with fMRI data and multi-objective optimization. Comput Methods Programs Biomed. 2020;196:105676.

232. Itani S, Rossignol M, Lecron F, Fortemps P. Towards interpretable machine learning models for diagnosis aid: a case study on attention deficit/hyperactivity disorder. PLoS One. 2019;14(4):e0215720.

233. Khoshnoud S, Nazari MA, Shamsi M. Functional brain dynamic analysis of ADHD and control children using nonlinear dynamical features of EEG signals. J Integr Neurosci. 2018;17(1):11–7.

234. Chikara RK, Ko LW. Neural activities classification of human inhibitory control using hierarchical model. Sensors (Basel). 2019;19(17):3791.

235. Dinkel PJ, Willmes K, Krinzinger H, Konrad K, Koten JW Jr. Diagnosing developmental dyscalculia on the basis of reliable single case FMRI methods: promises and limitations. PLoS One. 2013;8(12):e83722.

236. Zahia S, Garcia-Zapirain B, Saralegui I, Fernandez-Ruanova B. Dyslexia detection using 3D convolutional neural networks and functional

Chen *et al. BMC Medicine*   (2023) 21:241

Page 22 of 29

magnetic resonance imaging. Comput Methods Programs Biomed. 2020;197:105726.

237. Płoński P, Gradkowski W, Altarelli I, Monzalvo K, van Ermingen-Marbach M, Grande M, Heim S, Marchewka A, Bogorodzki P, Ramus F, et al. Multi-parameter machine learning approach to the neuroanatomical basis of developmental dyslexia. Hum Brain Mapp. 2017;38(2):900–8.

238. Martinez-Murcia FJ, Ortiz A, Gorriz JM, Ramirez J, Lopez-Abarejo PJ, Lopez-Zamora M, Luque JL. EEG connectivity analysis using denoising autoencoders for the detection of dyslexia. Int J Neural Syst. 2020;30(7):2050037.

239. Zainuddin AZA, Mansor W, Lee KY, Mahmoodin Z. Comparison of extreme learning machine and K-nearest neighbour performance in classifying EEG signal of normal, poor and capable dyslexic children. Annu Int Conf IEEE Eng Med Biol Soc. 2019;2019:4513–6.

240. Serrallach B, Groß C, Bernhofs V, Engelmann D, Benner J, Gündert N, Blatow M, Wengenroth M, Seitz A, Brunner M, et al. Neural biomarkers for dyslexia, ADHD, and ADD in the auditory cortex of children. Front Neurosci. 2016;10:324.

241. Cui Z, Xia Z, Su M, Shu H, Gong G. Disrupted white matter connectivity underlying developmental dyslexia: a machine learning approach. Hum Brain Mapp. 2016;37(4):1443–58.

242. García Chimeno Y, García Zapirain B, Saralegui Prieto I, Fernandez-Ruanova B. Automatic classification of dyslexic children by applying machine learning to fMRI images. Biomed Mater Eng. 2014;24(6):2995–3002.

243. Bailey S, Hoeft F, Aboud K, Cutting L. Anomalous gray matter patterns in specific reading comprehension deficit are independent of dyslexia. Ann Dyslexia. 2016;66(3):256–74.

244. Usman OL, Muniyandi RC, Omar K, Mohamad M. Gaussian smoothing and modified histogram normalization methods to improve neural-biomarker interpretations for dyslexia classification mechanism. PLoS One. 2021;16(2):e0245579.

245. Mascheretti S, Peruzzo D, Andreola C, Villa M, Ciceri T, Trezzi V, Marino C, Arrigoni F. Selecting the most relevant brain regions to classify children with developmental dyslexia and typical readers by using complex magnocellular stimuli and multiple kernel learning. Brain Sci. 2021;11(6):722.

246. Li A, Zalesky A, Yue W, Howes O, Yan H, Liu Y, Fan L, Whitaker KJ, Xu K, Rao G, et al. A neuroimaging biomarker for striatal dysfunction in schizophrenia. Nat Med. 2020;26(4):558–65.

247. Jo YT, Joo SW, Shon SH, Kim H, Kim Y, Lee J. Diagnosing schizophrenia with network analysis and a machine learning method. Int J Methods Psychiatr Res. 2020;29(1):e1818.

248. Yassin W, Nakatani H, Zhu Y, Kojima M, Owada K, Kuwabara H, Gonoi W, Aoki Y, Takao H, Natsubori T, et al. Machine-learning classification using neuroimaging data in schizophrenia, autism, ultra-high risk and first-episode psychosis. Transl Psychiatry. 2020;10(1):278.

249. Chen Z, Yan T, Wang E, Jiang H, Tang Y, Yu X, Zhang J, Liu C. Detecting abnormal brain regions in schizophrenia using structural MRI via machine learning. Comput Intell Neurosci. 2020;2020:6405930.

250. Jahmunah V, Lih OhS, Rajinikanth V, Ciaccio EJ, Hao Cheong K, Arunkumar N, Acharya UR. Automated detection of schizophrenia using nonlinear signal processing methods. Artif Intell Med. 2019;100:101698.

251. Wang L, Li X, Zhu Y, Lin B, Bo Q, Li F, Wang C. Discriminative analysis of symptom severity and ultra-high risk of schizophrenia using intrinsic functional connectivity. Int J Neural Syst. 2020;30(9):2050047.

252. Mikolas P, Hlinka J, Skoch A, Pitra Z, Frodl T, Spaniel F, Hajek T. Machine learning classification of first-episode schizophrenia spectrum disorders and controls using whole brain white matter fractional anisotropy. BMC Psychiatry. 2018;18(1):97.

253. Schwarz E, Doan NT, Pergola G, Westlye LT, Kaufmann T, Wolfers T, Brecheisen R, Quarto T, Ing AJ, Di Carlo P, et al. Reproducible grey matter patterns index a multivariate, global alteration of brain structure in schizophrenia and bipolar disorder. Transl Psychiatry. 2019;9(1):12.

254. Baradits M, Bitter I, Czobor P. Multivariate patterns of EEG microstate parameters and their role in the discrimination of patients with schizophrenia from healthy controls. Psychiatry Res. 2020;288:112938.

255. Kim J, Kim MY, Kwon H, Kim JW, Im WY, Lee SM, Kim K, Kim SJ. Feature optimization method for machine learning-based diagnosis of schizophrenia using magnetoencephalography. J Neurosci Methods. 2020;338:108688.

256. de Moura AM, Pinaya WHL, Gadelha A, Zugman A, Noto C, Cordeiro Q, Belangero SI, Jackowski AP, Bressan RA, Sato JR. Investigating brain structural patterns in first episode psychosis and schizophrenia using MRI and a machine learning approach. Psychiatry Res Neuroimaging. 2018;275:14–20.

257. Yamamoto M, Bagarinao E, Kushima I, Takahashi T, Sasabayashi D, Inada T, Suzuki M, Iidaka T, Ozaki N. Support vector machine-based classification of schizophrenia patients and healthy controls using structural magnetic resonance imaging from two independent sites. PLoS One. 2020;15(11):e0239615.

258. Zou H, Yang J. Dynamic thresholding networks for schizophrenia diagnosis. Artif Intell Med. 2019;96:25–32.

259. Rozycki M, Satterthwaite TD, Koutsouleris N, Erus G, Doshi J, Wolf DH, Fan Y, Gur RE, Gur RC, Meisenzahl EM, et al. Multisite machine learning analysis provides a robust structural imaging signature of schizophrenia detectable across diverse patient populations and within individuals. Schizophr Bull. 2018;44(5):1035–44.

260. Liang S, Deng W, Li X, Wang Q, Greenshaw AJ, Guo W, Kong X, Li M, Zhao L, Meng Y, et al. Aberrant posterior cingulate connectivity classify first-episode schizophrenia from controls: a machine learning study. Schizophr Res. 2020;220:187–93.

261. Alamian G, Pascarella A, Lajnef T, Knight L, Walters J, Singh KD, Jerbi K. Patient, interrupted: MEG oscillation dynamics reveal temporal dysconnectivity in schizophrenia. NeuroImage Clin. 2020;28:102485.

262. de Pierrefeu A, Löfstedt T, Laidi C, Hadj-Selem F, Bourgin J, Hajek T, Spaniel F, Kolenic M, Ciuciu P, Hamdani N, et al. Identifying a neuroanatomical signature of schizophrenia, reproducible across sites and stages, using machine learning with structured sparsity. Acta Psychiatr Scand. 2018;138(6):571–80.

263. Di Carlo P, Pergola G, Antonucci LA, Bonvino A, Mancini M, Quarto T, Rampino A, Popolizio T, Bertolino A, Blasi G. Multivariate patterns of gray matter volume in thalamic nuclei are associated with positive schizotypy in healthy individuals. Psychol Med. 2020;50(9):1501–9.

264. Bae Y, Kumarasamy K, Ali IM, Korfiatis P, Akkus Z, Erickson BJ. Differences between schizophrenic and normal subjects using network properties from fMRI. J Digit Imaging. 2018;31(2):252–61.

265. Yu Y, Shen H, Zeng LL, Ma Q, Hu D. Convergent and divergent functional connectivity patterns in schizophrenia and depression. PLoS One. 2013;8(7):e68250.

266. Dluhoš P, Schwarz D, Cahn W, van Haren N, Kahn R, Španiel F, Horáček J, Kašpárek T, Schnack H. Multi-center machine learning in imaging psychiatry: a meta-model approach. Neuroimage. 2017;155:10–24.

267. Lee J, Chon MW, Kim H, Rathi Y, Bouix S, Shenton ME, Kubicki M. Diagnostic value of structural and diffusion imaging measures in schizophrenia. NeuroImage Clin. 2018;18:467–74.

268. Antonucci LA, Penzel N, Pergola G, Kambeitz-Ilankovic L, Dwyer D, Kambeitz J, Haas SS, Passiatore R, Fazio L, Caforio G, et al. Multivariate classification of schizophrenia and its familial risk based on load-dependent attentional control brain functional connectivity. Neuropsychopharmacology. 2020;45(4):613–21.

269. Lu X, Yang Y, Wu F, Gao M, Xu Y, Zhang Y, Yao Y, Du X, Li C, Wu L, et al. Discriminative analysis of schizophrenia using support vector machine and recursive feature elimination on structural MRI images. Medicine (Baltim). 2016;95(30):e3973.

270. Chyzhyk D, Graña M, Öngür D, Shinn AK. Discrimination of schizophrenia auditory hallucinators by machine learning of resting-state functional MRI. Int J Neural Syst. 2015;25(3):1550007.

271. Jing R, Li P, Ding Z, Lin X, Zhao R, Shi L, Yan H, Liao J, Zhuo C, Lu L, et al. Machine learning identifies unaffected first-degree relatives with functional network patterns and cognitive impairment similar to those of schizophrenia patients. Hum Brain Mapp. 2019;40(13):3930–9.

272. Hua M, Peng Y, Zhou Y, Qin W, Yu C, Liang M. Disrupted pathways from limbic areas to thalamus in schizophrenia highlighted by whole-brain resting-state effective connectivity analysis. Prog Neuropsychopharmacol Biol Psychiatry. 2020;99:109837.

273. Gould IC, Shepherd AM, Laurens KR, Cairns MJ, Carr VJ, Green MJ. Multivariate neuroanatomical classification of cognitive subtypes in schizophrenia: a support vector machine learning approach. NeuroImage Clin. 2014;6:229–36.

274. Song H, Chen L, Gao R, Bogdan IIM, Yang J, Wang S, Dong W, Quan W, Dang W, Yu X. Automatic schizophrenic discrimination on fNIRS by

using complex brain network analysis and SVM. BMC Med Inform Decis Mak. 2017;17(Suppl 3):166.

275. Zhu Q, Huang J, Xu X. Non-negative discriminative brain functional connectivity for identifying schizophrenia on resting-state fMRI. Biomed Eng Online. 2018;17(1):32.

276. Zhao W, Guo S, Linli Z, Yang AC, Lin CP, Tsai SJ. Functional, anatomical, and morphological networks highlight the role of basal ganglia-thala-mus-cortex circuits in schizophrenia. Schizophr Bull. 2020;46(2):422–31.

277. Chin R, You AX, Meng F, Zhou J, Sim K. Recognition of schizophrenia with regularized support vector machine and sequential region of interest selection using structural magnetic resonance imaging. Sci Rep. 2018;8(1):13858.

278. Iwabuchi SJ, Palaniyappan L. Abnormalities in the effective con-nectivity of visuothalamic circuitry in schizophrenia. Psychol Med. 2017;47(7):1300–10.

279. Winterburn JL, Voineskos AN, Devenyi GA, Plitman E, de la Fuente-Sand-oval C, Bhagwat N, Graff-Guerrero A, Knight J, Chakravarty MM. Can we accurately classify schizophrenia patients from healthy controls using magnetic resonance imaging and machine learning? A multi-method and multi-dataset study. Schizophr Res. 2019;214:3–10.

280. Chen H, Uddin LQ, Duan X, Zheng J, Long Z, Zhang Y, Guo X, Zhang Y, Zhao J, Chen H. Shared atypical default mode and salience network functional connectivity between autism and schizophrenia. Autism Res. 2017;10(11):1776–86.

281. Zeng LL, Wang H, Hu P, Yang B, Pu W, Shen H, Chen X, Liu Z, Yin H, Tan Q, et al. Multi-site diagnostic classification of schizophrenia using discri-minant deep learning with functional connectivity MRI. EBioMedicine. 2018;30:74–85.

282. Supekar K, Cai W, Krishnadas R, Palaniyappan L, Menon V. Dysregulated brain dynamics in a triple-network saliency model of schizophrenia and its relation to psychosis. Biol Psychiatry. 2019;85(1):60–9.

283. Mastrovito D, Hanson C, Hanson SJ. Differences in atypical resting-state effective connectivity distinguish autism from schizophrenia. NeuroIm-age Clin. 2018;18:367–76.

284. Liu Y, Zhang Y, Lv L, Wu R, Zhao J, Guo W. Abnormal neural activity as a potential biomarker for drug-naive first-episode adolescent-onset schizophrenia with coherence regional homogeneity and support vec-tor machine analyses. Schizophr Res. 2018;192:408–15.

285. Cui LB, Liu L, Wang HN, Wang LX, Guo F, Xi YB, Liu TT, Li C, Tian P, Liu K, et al. Disease definition for schizophrenia by functional connectivity using radiomics strategy. Schizophr Bull. 2018;44(5):1053–9.

286. Pinaya WHL, Mechelli A, Sato JR. Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: a large-scale multi-sample study. Hum Brain Mapp. 2019;40(3):944–54.

287. Chen X, Liu C, He H, Chang X, Jiang Y, Li Y, Duan M, Li J, Luo C, Yao D. Transdiagnostic differences in the resting-state functional connectiv-ity of the prefrontal cortex in depression and schizophrenia. J Affect Disord. 2017;217:118–24.

288. Schnack HG, Nieuwenhuis M, van Haren NE, Abramovic L, Scheewe TW, Brouwer RM, Hulshoff Pol HE, Kahn RS. Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. Neuroimage. 2014;84:299–306.

289. Cheng H, Newman S, Goñi J, Kent JS, Howell J, Bolbecker A, Puce A, O'Donnell BF, Hetrick WP. Nodal centrality of functional network in the differentiation of schizophrenia. Schizophr Res. 2015;168(1–2):345–52.

290. Mikolas P, Melicher T, Skoch A, Matejka M, Slovakova A, Bakstein E, Hajek T, Spaniel F. Connectivity of the anterior insula differenti-ates participants with first-episode schizophrenia spectrum disorders from controls: a machine-learning study. Psychol Med. 2016;46(13):2695–704.

291. Venkataraman A, Whitford TJ, Westin CF, Golland P, Kubicki M. Whole brain resting state functional connectivity abnormalities in schizophre-nia. Schizophr Res. 2012;139(1–3):7–12.

292. Wang S, Zhang Y, Lv L, Wu R, Fan X, Zhao J, Guo W. Abnormal regional homogeneity as a potential imaging biomarker for adolescent-onset schizophrenia: a resting-state fMRI study and support vector machine analysis. Schizophr Res. 2018;192:179–84.

293. Cabral C, Kambeitz-Ilankovic L, Kambeitz J, Calhoun VD, Dwyer DB, von Saldern S, Urquijo MF, Falkai P, Koutsouleris N. Classifying schizophrenia using multimodal multivariate pattern recognition analysis: evaluating

the impact of individual clinical profiles on the neurodiagnostic perfor-mance. Schizophr Bull. 2016;42 Suppl 1(Suppl 1):S110-117.

294. Ebdrup BH, Axelsen MC, Bak N, Fagerlund B, Oranje B, Raghava JM, Nielsen M, Rostrup E, Hansen LK, Glenthøj BY. Accuracy of diagnostic classification algorithms using cognitive-, electrophysiological-, and neuroanatomical data in antipsychotic-naïve schizophrenia patients. Psychol Med. 2019;49(16):2754–63.

295. Pinaya WH, Gadelha A, Doyle OM, Noto C, Zugman A, Cordeiro Q, Jackowski AP, Bressan RA, Sato JR. Using deep belief network modelling to characterize differences in brain morphometry in schizophrenia. Sci Rep. 2016;6:38897.

296. Kim J, Calhoun VD, Shim E, Lee JH. Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. Neuro-image. 2016;124(Pt A):127–46.

297. Viviano JD, Buchanan RW, Calarco N, Gold JM, Foussias G, Bhagwat N, Stefanik L, Hawco C, DeRosse P, Argyelan M, et al. Resting-state con-nectivity biomarkers of cognitive performance and social function in individuals with schizophrenia spectrum disorder and healthy control subjects. Biol Psychiatry. 2018;84(9):665–74.

298. Arbabshirani MR, Castro E, Calhoun VD. Accurate classification of schizo-phrenia patients based on novel resting-state fMRI features. Annu Int Conf IEEE Eng Med Biol Soc. 2014;2014:6691–4.

299. Yan W, Calhoun V, Song M, Cui Y, Yan H, Liu S, Fan L, Zuo N, Yang Z, Xu K, et al. Discriminating schizophrenia using recurrent neural net-work applied on time courses of multi-site FMRI data. EBioMedicine. 2019;47:543–52.

300. Orban P, Dansereau C, Desbois L, Mongeau-Pérusse V, Giguère C, Nguyen H, Mendrek A, Stip E, Bellec P. Multisite generalizability of schizophrenia diagnosis classification based on functional brain con-nectivity. Schizophr Res. 2018;192:167–71.

301. Yu Y, Shen H, Zhang H, Zeng LL, Xue Z, Hu D. Functional connectivity-based signatures of schizophrenia revealed by multiclass pattern analy-sis of resting-state fMRI from schizophrenic patients and their healthy siblings. Biomed Eng Online. 2013;12:10.

302. Dillon K, Calhoun V, Wang YP. A robust sparse-modeling framework for estimating schizophrenia biomarkers from fMRI. J Neurosci Methods. 2017;276:46–55.

303. Pergola G, Trizio S, Di Carlo P, Taurisano P, Mancini M, Amoroso N, Nettis MA, Andriola I, Caforio G, Popolizio T, et al. Grey matter volume patterns in thalamic nuclei are associated with familial risk for schizophrenia. Schizophr Res. 2017;180:13–20.

304. Shen H, Wang L, Liu Y, Hu D. Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimen-sional embedding of fMRI. Neuroimage. 2010;49(4):3110–21.

305. Chu WL, Huang MW, Jian BL, Hsu CY, Cheng KS. A Correlative clas-sification study of schizophrenic patients with results of clinical evaluation and structural magnetic resonance images. Behav Neurol. 2016;2016:7849526.

306. Masychev K, Ciprian C, Ravan M, Reilly JP, MacCrimmon D. Advanced signal processing methods for characterization of schizophrenia. IEEE Trans Biomed Eng. 2021;68(4):1123–30.

307. Korda AI, Ruef A, Neufang S, Davatzikos C, Borgwardt S, Meisenzahl EM, Koutsouleris N. Identification of voxel-based texture abnormalities as new biomarkers for schizophrenia and major depressive patients using layer-wise relevance propagation on deep learning decisions. Psychia-try Res Neuroimaging. 2021;313:111303.

308. Pina-Camacho L, Garcia-Prieto J, Parellada M, Castro-Fornieles J, Gon-zalez-Pinto AM, Bombin I, Graell M, Paya B, Rapado-Castro M, Janssen J, et al. Predictors of schizophrenia spectrum disorders in early-onset first episodes of psychosis: a support vector machine model. Eur Child Adolesc Psychiatry. 2015;24(4):427–40.

309. Castro E, Martínez-Ramón M, Pearlson G, Sui J, Calhoun VD. Character-ization of groups using composite kernels and multi-source fMRI analy-sis data: application to schizophrenia. Neuroimage. 2011;58(2):526–36.

310. Koch SP, Hägele C, Haynes JD, Heinz A, Schlagenhauf F, Sterzer P. Diag-nostic classification of schizophrenia patients on the basis of regional reward-related FMRI signal patterns. PLoS One. 2015;10(3):e0119089.

311. Castro E, Gupta CN, Martínez-Ramón M, Calhoun VD, Arbabshirani MR, Turner J. Identification of patterns of gray matter abnormalities in

Chen *et al. BMC Medicine*   (2023) 21:241

Page 24 of 29

schizophrenia using source-based morphometry and bagging. Annu Int Conf IEEE Eng Med Biol Soc. 2014;2014:1513–6.

312. Arribas JI, Calhoun VD, Adali T. Automatic Bayesian classification of healthy controls, bipolar disorder, and schizophrenia using intrinsic connectivity maps from FMRI data. IEEE Trans Biomed Eng. 2010;57(12):2850–60.

313. Yoon JH, Nguyen DV, McVay LM, Deramo P, Minzenberg MJ, Ragland JD, Niendham T, Solomon M, Carter CS. Automated classification of fMRI during cognitive control identifies more severely disorganized subjects with schizophrenia. Schizophr Res. 2012;135(1–3):28–33.

314. Chyzhyk D, Savio A, Graña M. Computer aided diagnosis of schizophrenia on resting state fMRI data by ensembles of ELM. Neural Netw. 2015;68:23–33.

315. Wang P, Verma R. On classifying disease-induced patterns in the brain using diffusion tensor images. Med Image Comput Comput Assist Interv. 2008;11(Pt 1):908–16.

316. Ince NF, Goksu F, Pellizzer G, Tewfik A, Stephane M. Selection of spectro-temporal patterns in multichannel MEG with support vector machines for schizophrenia classification. Annu Int Conf IEEE Eng Med Biol Soc. 2008;2008:3554–7.

317. Sveinsson JR, Benediktsson JA, Stefansson SB, Davidsson K. Parallel principal component neural networks for classification of event-related potential waveforms. Med Eng Phys. 1997;19(1):15–20.

318. Calhas D, Romero E, Henriques R. On the use of pairwise distance learning for brain signal classification with limited observations. Artif Intell Med. 2020;105:101852.

319. Calhoun VD, Maciejewski PK, Pearlson GD. Temporal lobe and "Default" hemodynamic brain modes discriminate between schizophrenia and bipolar disorder. Hum Brain Mapp. 2008;29(11):1265–75.

320. Neuhaus AH, Popescu FC, Bates JA, Goldberg TE, Malhotra AK. Single-subject classification of schizophrenia using event-related potentials obtained during auditory and visual oddball paradigms. Eur Arch Psychiatry Clin Neurosci. 2013;263(3):241–7.

321. Xu T, Stephane M, Parhi KK. Abnormal neural oscillations in schizophrenia assessed by spectral power ratio of MEG during word processing. IEEE Trans Neural Syst Rehabil Eng. 2016;24(11):1148–58.

322. Ravan M, MacCrimmon D, Hasey G, Reilly JP, Khodayari-Rostamabad A. A machine learning approach using P300 responses to investigate effect of clozapine therapy. Annu Int Conf IEEE Eng Med Biol Soc. 2012;2012:5911–4.

323. Castellani U, Rossato E, Murino V, Bellani M, Rambaldelli G, Perlini C, Tomelleri L, Tansella M, Brambilla P. Classification of schizophrenia using feature-based morphometry. J Neural Transm (Vienna, Austria : 1996). 2012;119(3):395–404.

324. Khare S, Bajaj V, Siuly S, Sinha PG. Classification of schizophrenia patients through empirical wavelet transformation using electroencephalogram signals. In., edn.; 2020: 1.1–1.26.

325. Yan W, Zhao M, Fu Z, Pearlson GD, Sui J, Calhoun VD. Mapping relationships among schizophrenia, bipolar and schizoaffective disorders: a deep classification and clustering framework using fMRI time series. Schizophr Res. 2022;245:141–50.

326. Du Y, Hao H, Wang S, Pearlson GD, Calhoun VD. Identifying commonality and specificity across psychosis sub-groups via classification based on features from dynamic connectivity analysis. NeuroImage Clin. 2020;27:102284.

327. Oh K, Kim W, Shen G, Piao Y, Kang NI, Oh IS, Chung YC. Classification of schizophrenia and normal controls using 3D convolutional neural network and outcome visualization. Schizophr Res. 2019;212:186–95.

328. Guo Y, Qiu J, Lu W. Support vector machine-based schizophrenia classification using morphological information from amygdaloid and hippocampal subregions. Brain Sci. 2020;10(8):562.

329. Qureshi MNI, Oh J, Cho D, Jo HJ, Lee B. Multimodal discrimination of schizophrenia using hybrid weighted feature concatenation of brain functional connectivity and anatomical features with an extreme learning machine. Front Neuroinform. 2017;11:59.

330. Gallos IK, Gkiatis K, Matsopoulos GK, Siettos C. ISOMAP and machine learning algorithms for the construction of embedded functional connectivity networks of anatomically separated brain regions from resting state fMRI data of patients with Schizophrenia. AIMS Neurosci. 2021;8(2):295–321.

331. Wang T, Bezerianos A, Cichocki A, Li J. Multikernel Capsule Network for Schizophrenia Identification. IEEE Trans Cybern. 2022;52(6):4741–50.

332. Zang J, Huang Y, Kong L, Lei B, Ke P, Li H, Zhou J, Xiong D, Li G, Chen J, et al. Effects of brain atlases and machine learning methods on the discrimination of schizophrenia patients: a multimodal MRI study. Front Neurosci. 2021;15:697168.

333. Hu M, Qian X, Liu S, Koh AJ, Sim K, Jiang X, Guan C, Zhou JH. Structural and diffusion MRI based schizophrenia classification using 2D pre-trained and 3D naive Convolutional Neural Networks. Schizophr Res. 2022;243:330–41.

334. Kim K, Duc NT, Choi M, Lee B. EEG microstate features for schizophrenia classification. PLoS One. 2021;16(5):e0251842.

335. Salvador R, Canales-Rodríguez E, Guerrero-Pedraza A, Sarró S, Tordesillas-Gutiérrez D, Maristany T, Crespo-Facorro B, McKenna P, Pomarol-Clotet E. Multimodal integration of brain images for MRI-based diagnosis in schizophrenia. Front Neurosci. 2019;13:1203.

336. Chou PH, Yao YH, Zheng RX, Liou YL, Liu TT, Lane HY, Yang AC, Wang SC. Deep neural network to differentiate brain activity between patients with first-episode schizophrenia and healthy individuals: a multi-channel near infrared spectroscopy study. Front Psych. 2021;12:655292.

337. Li Z, Li W, Wei Y, Gui G, Zhang R, Liu H, Chen Y, Jiang Y. Deep learning based automatic diagnosis of first-episode psychosis, bipolar disorder and healthy controls. Comput Med Imaging Graph. 2021;89:101882.

338. Yang H, Liu J, Sui J, Pearlson G, Calhoun VD. A hybrid machine learning method for fusing fMRI and genetic data: combining both improves classification of schizophrenia. Front Hum Neurosci. 2010;4:192.

339. Kalmady SV, Greiner R, Agrawal R, Shivakumar V, Narayanaswamy JC, Brown MRG, Greenshaw AJ, Dursun SM, Venkatasubramanian G. Towards artificial intelligence in mental health by improving schizophrenia prediction with multiple brain parcellation ensemble-learning. NPJ Schizophr. 2019;5(1):2.

340. Liu W, Zhang X, Qiao Y, Cai Y, Yin H, Zheng M, Zhu Y, Wang H. Functional connectivity combined with a machine learning algorithm can classify high-risk first-degree relatives of patients with schizophrenia and identify correlates of cognitive impairments. Front Neurosci. 2020;14:577568.

341. Bansal R, Staib LH, Laine AF, Hao X, Xu D, Liu J, Weissman M. Anatomical brain images alone can accurately diagnose chronic neuropsychiatric illnesses. PLoS One. 2012;7(12):e50698.

342. Li F, Wang J, Liao Y, Yi C, Jiang Y, Si Y, Peng W, Yao D, Zhang Y, Dong W, et al. Differentiation of schizophrenia by combining the spatial EEG brain network patterns of rest and task P300. IEEE Trans Neural Syst Rehabil Eng. 2019;27(4):594–602.

343. Zhao Z, Li J, Niu Y, Wang C, Zhao J, Yuan Q, Ren Q, Xu Y, Yu Y. Classification of schizophrenia by combination of brain effective and functional connectivity. Front Neurosci. 2021;15:651439.

344. Park SM, Jeong B, Oh DY, Choi CH, Jung HY, Lee JY, Lee D, Choi JS. Identification of major psychiatric disorders from resting-state electroencephalography using a machine learning approach. Front Psych. 2021;12:707581.

345. Lei D, Pinaya WHL, Young J, van Amelsvoort T, Marcelis M, Donohoe G, Mothersill DO, Corvin A, Vieira S, Huang X, et al. Integrating machining learning and multimodal neuroimaging to detect schizophrenia at the level of the individual. Hum Brain Mapp. 2020;41(5):1119–35.

346. Kim JY, Lee HS, Lee SH. EEG source network for the diagnosis of schizophrenia and the identification of subtypes based on symptom severity-a machine learning approach. J Clin Med. 2020;9(12):3934.

347. Shi D, Li Y, Zhang H, Yao X, Wang S, Wang G, Ren K. Machine learning of schizophrenia detection with structural and functional neuroimaging. Dis Markers. 2021;2021:9963824.

348. Shalbaf A, Bagherzadeh S, Maghsoudi A. Transfer learning with deep convolutional neural network for automated detection of schizophrenia from EEG signals. Phys Eng Sci Med. 2020;43(4):1229–39.

349. Janousova E, Montana G, Kasparek T, Schwarz D. Supervised, multivariate, whole-brain reduction did not help to achieve high classification performance in schizophrenia research. Front Neurosci. 2016;10:392.

350. Lieslehto J, Jääskeläinen E, Kiviniemi V, Haapea M, Jones PB, Murray GK, Veijola J, Dannlowski U, Grotegerd D, Meinert S, et al. The progression of disorder-specific brain pattern expression in schizophrenia over 9 years. NPJ Schizophr. 2021;7(1):32.

Chen *et al. BMC Medicine*   (2023) 21:241

Page 25 of 29

351. Ke PF, Xiong DS, Li JH, Pan ZL, Zhou J, Li SJ, Song J, Chen XY, Li GX, Chen J, et al. An integrated machine learning framework for a discriminative analysis of schizophrenia using multi-biological data. Sci Rep. 2021;11(1):14636.

352. Gheiratmand M, Rish I, Cecchi GA, Brown MRG, Greiner R, Polosecki PI, Bashivan P, Greenshaw AJ, Ramasubbu R, Dursun SM. Learning stable and predictive network-based patterns of schizophrenia and its clinical symptoms. NPJ Schizophr. 2017;3:22.

353. Chen J, Li X, Calhoun VD, Turner JA, van Erp TGM, Wang L, Andreassen OA, Agartz I, Westlye LT, Jönsson E, et al. Sparse deep neural networks on imaging genetics for schizophrenia case-control classification. Hum Brain Mapp. 2021;42(8):2556–68.

354. Oh J, Oh BL, Lee KU, Chae JH, Yun K. Identifying schizophrenia using structural MRI with a deep learning algorithm. Front Psych. 2020;11:16.

355. Singh K, Singh S, Malhotra J. Spectral features based convolutional neural network for accurate and prompt identification of schizophrenic patients. Proc Inst Mech Eng H. 2021;235(2):167–84.

356. Sun J, Cao R, Zhou M, Hussain W, Wang B, Xue J, Xiang J. A hybrid deep neural network for classification of schizophrenia using EEG Data. Sci Rep. 2021;11(1):4706.

357. Li YJ, Fan FY. Classification of Schizophrenia and Depression by EEG with ANNs. Conf Proc IEEE Eng Med Biol Soc. 2005;2005:2679–82.

358. Johannesen JK, Bi J, Jiang R, Kenney JG, Chen CA. Machine learning identification of EEG features predicting working memory performance in schizophrenia and healthy adults. Neuropsychiatr Electrophysiol. 2016;2:3.

359. Jin K, Xu D, Shen Z, Feng G, Zhao Z, Lu J, Lyu H, Pan F, Shang D, Chen J, et al. Distinguishing hypochondriasis and schizophrenia using regional homogeneity: a resting-state fMRI study and support vector machine analysis. Acta Neuropsychiatr. 2021;33(4):182–90.

360. Siuly S, Khare SK, Bajaj V, Wang H, Zhang Y. A Computerized method for automatic detection of schizophrenia using EEG signals. IEEE Trans Neural Syst Rehabil Eng. 2020;28(11):2390–400.

361. Li H, Cui L, Cao L, Zhang Y, Liu Y, Deng W, Zhou W. Identification of bipolar disorder using a combination of multimodality magnetic resonance imaging and machine learning techniques. BMC Psychiatry. 2020;20(1):488.

362. Linke JO, Adleman NE, Sarlls J, Ross A, Perlstein S, Frank HR, Towbin KE, Pine DS, Leibenluft E, Brotman MA. White matter microstructure in pediatric bipolar disorder and disruptive mood dysregulation disorder. J Am Acad Child Adolesc Psychiatry. 2020;59(10):1135–45.

363. Squarcina L, Dagnew TM, Rivolta MW, Bellani M, Sassi R, Brambilla P. Automated cortical thickness and skewness feature selection in bipolar disorder using a semi-supervised learning method. J Affect Disord. 2019;256:416–23.

364. Matsuo K, Harada K, Fujita Y, Okamoto Y, Ota M, Narita H, Mwangi B, Gutierrez CA, Okada G, Takamura M, et al. Distinctive neuroanatomical substrates for depression in bipolar disorder versus major depressive disorder. Cereb Cortex. 2019;29(1):202–14.

365. Frangou S, Dima D, Jogia J. Towards person-centered neuroimaging markers for resilience and vulnerability in Bipolar Disorder. Neuroimage. 2017;145(Pt B):230–7.

366. Doan NT, Kaufmann T, Bettella F, Jørgensen KN, Brandt CL, Moberget T, Alnæs D, Douaud G, Duff E, Djurovic S, et al. Distinct multivariate brain morphological patterns and their added predictive value with cognitive and polygenic risk scores in mental disorders. NeuroImage Clin. 2017;15:719–31.

367. Calhoun VD, Maciejewski PK, Pearlson GD, Kiehl KA. Temporal lobe and "default" hemodynamic brain modes discriminate between schizophrenia and bipolar disorder. Hum Brain Mapp. 2008;29(11):1265–75.

368. Reavis EA, Lee J, Wynn JK, Engel SA, Cohen MS, Nuechterlein KH, Glahn DC, Altshuler LL, Green MF. Assessing neural tuning for object perception in schizophrenia and bipolar disorder with multivariate pattern analysis of fMRI data. NeuroImage Clinical. 2017;16:491–7.

369. Mwangi B, Wu MJ, Bauer IE, Modi H, Zeni CP, Zunta-Soares GB, Hasan KM, Soares JC. Predictive classification of pediatric bipolar disorder using atlas-based diffusion weighted imaging and support vector machines. Psychiatry Res. 2015;234(2):265–71.

370. Mwangi B, Spiker D, Zunta-Soares GB, Soares JC. Prediction of pediatric bipolar disorder using neuroanatomical signatures of the amygdala. Bipolar Disord. 2014;16(7):713–21.

371. Costafreda SG, Fu CH, Picchioni M, Toulopoulou T, McDonald C, Kravariti E, Walshe M, Prata D, Murray RM, McGuire PK. Pattern of neural responses to verbal fluency shows diagnostic specificity for schizophrenia and bipolar disorder. BMC Psychiatry. 2011;11:18.

372. Rashid B, Arbabshirani MR, Damaraju E, Cetin MS, Miller R, Pearlson GD, Calhoun VD. Classification of schizophrenia and bipolar patients using static and dynamic resting-state fMRI brain connectivity. Neuroimage. 2016;134:645–57.

373. Kaufmann T, Alnæs D, Brandt CL, Doan NT, Kauppi K, Bettella F, Lagerberg TV, Berg AO, Djurovic S, Agartz I, et al. Task modulations and clinical manifestations in the brain functional connectome in 1615 fMRI datasets. Neuroimage. 2017;147:243–52.

374. Mourão-Miranda J, Almeida JR, Hassel S, de Oliveira L, Versace A, Marquand AF, Sato JR, Brammer M, Phillips ML. Pattern recognition analyses of brain activation elicited by happy and neutral faces in unipolar and bipolar depression. Bipolar Disord. 2012;14(4):451–60.

375. Wu MJ, Mwangi B, Bauer IE, Passos IC, Sanches M, Zunta-Soares GB, Meyer TD, Hasan KM, Soares JC. Identification and individualized prediction of clinical phenotypes in bipolar disorders using neurocognitive data, neuroimaging scans and machine learning. Neuroimage. 2017;145(Pt B):254–64.

376. Bürger C, Redlich R, Grotegerd D, Meinert S, Dohm K, Schneider I, Zaremba D, Förster K, Alferink J, Bölte J, et al. Differential abnormal pattern of anterior cingulate gyrus activation in unipolar and bipolar depression: an fMRI and pattern classification approach. Neuropsychopharmacology. 2017;42(7):1399–408.

377. Anticevic A, Cole MW, Repovs G, Murray JD, Brumbaugh MS, Winkler AM, Savic A, Krystal JH, Pearlson GD, Glahn DC. Characterizing thalamo-cortical disturbances in schizophrenia and bipolar illness. Cereb Cortex. 2014;24(12):3116–30.

378. Besga A, Termenon M, Graña M, Echeveste J, Pérez JM, Gonzalez-Pinto A. Discovering Alzheimer's disease and bipolar disorder white matter effects building computer aided diagnostic systems on brain diffusion tensor imaging features. Neurosci Lett. 2012;520(1):71–6.

379. Chen Y, Storrs J, Tan L, Mazlack LJ, Lee JH, Lu LJ. Detecting brain structural changes as biomarker from magnetic resonance images using a local feature based SVM approach. J Neurosci Methods. 2014;221:22–31.

380. Xi C, Lai J, Du Y, Ng CH, Jiang J, Wu L, Zhang P, Xu Y, Hu S. Abnormal functional connectivity within the reward network: a potential neuroimaging endophenotype of bipolar disorder. J Affect Disord. 2021;280(Pt B):49–56.

381. Teng S, Lu CF, Wang PS, Hung CI, Li CT, Tu PC, Su TP, Wu YT. Classification of bipolar disorder using basal-ganglia-related functional connectivity in the resting state. Annu Int Conf IEEE Eng Med Biol Soc. 2013;2013:1057–60.

382. Bansal R, Staib LH, Laine AF, Hao X, Xu D, Liu J, Weissman M, Peterson BS. Anatomical brain images alone can accurately diagnose chronic neuropsychiatric illnesses. PLoS One. 2012;7(12):e50698.

383. Nazhvani AD, Boostani R, Afrasiabi S, Sadatnezhad K. Classification of ADHD and BMD patients using visual evoked potential. Clin Neurol Neurosurg. 2013;115(11):2329–35.

384. Chen YL, Tu PC, Huang TH, Bai YM, Su TP, Chen MH, Wu YT. Using minimal-redundant and maximal-relevant whole-brain functional connectivity to classify bipolar disorder. Front Neurosci. 2020;14:563368.

385. Mwangi B, Wu MJ, Cao B, Passos IC, Lavagnino L, Keser Z, Zunta-Soares GB, Hasan KM, Kapczinski F, Soares JC. Individualized prediction and clinical staging of bipolar disorders using neuroanatomical biomarkers. Biol Psychiatry Cogn Neurosci Neuroimaging. 2016;1(2):186–94.

386. Yamashita A, Sakai Y, Yamada T, Yahata N, Kunimatsu A, Okada N, Itahashi T, Hashimoto R, Mizuta H, Ichikawa N, et al. Generalizable brain network markers of major depressive disorder across multiple imaging sites. PLoS Biol. 2020;18(12):e3000966.

387. Maglanoc LA, Kaufmann T, Jonassen R, Hilland E, Beck D, Landrø NI, Westlye LT. Multimodal fusion of structural and functional brain imaging in depression using linked independent component analysis. Hum Brain Mapp. 2020;41(1):241–55.

388. Shim M, Jin MJ, Im CH, Lee SH. Machine-learning-based classification between post-traumatic stress disorder and major depressive disorder using P300 features. NeuroImage Clinical. 2019;24:102001.

389. Chun JY, Sendi MSE, Sui J, Zhi D, Calhoun VD. Visualizing functional network connectivity difference between healthy control and major

Chen *et al. BMC Medicine*   (2023) 21:241

Page 26 of 29

depressive disorder using an explainable machine-learning method. Annu Int Conf IEEE Eng Med Biol Soc. 2020;2020:1424–7.

390. Uyulan C, Ergüzel TT, Unubol H, Cebi M, Sayar GH, Nezhad Asad M, Tarhan N. Major depressive disorder classification based on different convolutional neural network models: deep learning approach. Clin EEG Neurosci. 2021;52(1):38–51.

391. Shi Y, Zhang L, Wang Z, Lu X, Wang T, Zhou D, Zhang Z. Multivariate machine learning analyses in identification of major depressive disorder using resting-state functional connectivity: a multicentral study. ACS Chem Neurosci. 2021;12(15):2878–86.

392. Mahato S, Goyal N, Ram D, Paul S. Detection of depression and scaling of severity using six channel EEG data. J Med Syst. 2020;44(7):118.

393. Guo H, Cao X, Liu Z, Li H, Chen J, Zhang K. Machine learning classifier using abnormal brain network topological metrics in major depressive disorder. NeuroReport. 2012;23(17):1006–11.

394. Guo H, Li Y, Mensah GK, Xu Y, Chen J, Xiang J, Chen D. Resting-State Functional Network Scale Effects and Statistical Significance-Based Feature Selection in Machine Learning Classification. Comput Math Methods Med. 2019;2019:9108108.

395. Patel MJ, Andreescu C, Price JC, Edelman KL, Reynolds CF 3rd, Aizenstein HJ. Machine learning approaches for integrating clinical and imaging features in late-life depression classification and response prediction. Int J Geriatr Psychiatry. 2015;30(10):1056–67.

396. Yang J, Zhang M, Ahn H, Zhang Q, Jin TB, Li I, Nemesure M, Joshi N, Jiang H, Miller JM, et al. Development and evaluation of a multimodal marker of major depressive disorder. Hum Brain Mapp. 2018;39(11):4420–39.

397. Ahmadlou M, Adeli H, Adeli A. Fractality analysis of frontal brain in major depressive disorder. Int J Psychophysiol. 2012;85(2):206–11.

398. Sacchet MD, Livermore EE, Iglesias JE, Glover GH, Gotlib IH. Subcortical volumes differentiate major depressive disorder, bipolar disorder, and remitted major depressive disorder. J Psychiatr Res. 2015;68:91–8.

399. Ramasubbu R, Brown EC, Marcil LD, Talai AS, Forkert ND. Automatic classification of major depression disorder using arterial spin labeling MRI perfusion measurements. Psychiatry Clin Neurosci. 2019;73(8):486–93.

400. Li H, Song S, Wang D, Tan Z, Lian Z, Wang Y, Zhou X, Pan C. Individualized diagnosis of major depressive disorder via multivariate pattern analysis of thalamic sMRI features. BMC Psychiatry. 2021;21(1):415.

401. Zhong X, Shi H, Ming Q, Dong D, Zhang X, Zeng LL, Yao S. Whole-brain resting-state functional connectivity identified major depressive disorder: a multivariate pattern analysis in two independent samples. J Affect Disord. 2017;218:346–52.

402. Liu W, Zhang C, Wang X, Xu J, Chang Y, Ristaniemi T, Cong F. Functional connectivity of major depression disorder using ongoing EEG during music perception. Clin Neurophysiol. 2020;131(10):2413–22.

403. Liao SC, Wu CT, Huang HC, Cheng WT, Liu YH. Major depression detection from EEG signals using kernel eigen-filter-bank common spatial patterns. Sensors (Basel). 2017;17(6):1385.

404. Rosa MJ, Portugal L, Hahn T, Fallgatter AJ, Garrido MI, Shawe-Taylor J, Mourao-Miranda J. Sparse network-based models for patient classification using fMRI. Neuroimage. 2015;105:493–506.

405. Schnyer DM, Clasen PC, Gonzalez C, Beevers CG. Evaluating the diagnostic utility of applying a machine learning algorithm to diffusion tensor MRI measures in individuals with major depressive disorder. Psychiatry Res Neuroimaging. 2017;264:1–9.

406. Johnston BA, Steele JD, Tolomeo S, Christmas D, Matthews K. Structural MRI-Based Predictions in Patients with Treatment-Refractory Depression (TRD). PLoS One. 2015;10(7):e0132958.

407. Liu F, Guo W, Yu D, Gao Q, Gao K, Xue Z, Du H, Zhang J, Tan C, Liu Z, et al. Classification of different therapeutic responses of major depressive disorder with multivariate pattern analysis method based on structural MR scans. PLoS One. 2012;7(7):e40968.

408. Wei M, Qin J, Yan R, Li H, Yao Z, Lu Q. Identifying major depressive disorder using Hurst exponent of resting-state brain networks. Psychiatry Res. 2013;214(3):306–12.

409. Chu SH, Lenglet C, Schreiner MW, Klimes-Dougan B, Cullen K, Parhi KK. Anatomical biomarkers for adolescent major depressive disorder from diffusion weighted imaging using SVM classifier. Annu Int Conf IEEE Eng Med Biol Soc. 2018;2018:2740–3.

410. Guo H, Qin M, Chen J, Xu Y, Xiang J. Machine-learning classifier for patients with major depressive disorder: multifeature approach based

on a high-order minimum spanning tree functional brain network. Comput Math Methods Med. 2017;2017:4820935.

411. Zeng LL, Shen H, Liu L, Hu D. Unsupervised classification of major depression using functional connectivity MRI. Hum Brain Mapp. 2014;35(4):1630–41.

412. Fang P, Zeng LL, Shen H, Wang L, Li B, Liu L, Hu D. Increased cortical-limbic anatomical network connectivity in major depression revealed by diffusion tensor imaging. PLoS One. 2012;7(9):e45972.

413. Tan W, Liu Z, Xi C, Deng M, Long Y, Palaniyappan L, Yang J. Decreased integration of the frontoparietal network during a working memory task in major depressive disorder. Aust N Z J Psychiatry. 2021;55(6):577–87.

414. Cao L, Guo S, Xue Z, Hu Y, Liu H, Mwansisya TE, Pu W, Yang B, Liu C, Feng J, et al. Aberrant functional connectivity for diagnosis of major depressive disorder: a discriminant analysis. Psychiatry Clin Neurosci. 2014;68(2):110–9.

415. Zeng LL, Shen H, Liu L, Wang L, Li B, Fang P, Zhou Z, Li Y, Hu D. Identifying major depression using whole-brain functional connectivity: a multivariate pattern analysis. Brain. 2012;135(Pt 5):1498–507.

416. Ramasubbu R, Brown MR, Cortese F, Gaxiola I, Goodyear B, Greenshaw AJ, Dursun SM, Greiner R. Accuracy of automated classification of major depressive disorder as a function of symptom severity. NeuroImage Clin. 2016;12:320–31.

417. Qiu L, Huang X, Zhang J, Wang Y, Kuang W, Li J, Wang X, Wang L, Yang X, Lui S, et al. Characterization of major depressive disorder using a multiparametric classification approach based on high resolution structural images. J Psychiatry Neurosci. 2014;39(2):78–86.

418. Mwangi B, Ebmeier KP, Matthews K, Steele JD. Multi-centre diagnostic classification of individual structural neuroimaging scans from patients with major depressive disorder. Brain. 2012;135(Pt 5):1508–21.

419. Zhao J, Huang J, Zhi D, Yan W, Ma X, Yang X, Li X, Ke Q, Jiang T, Calhoun VD, et al. Functional network connectivity (FNC)-based generative adversarial network (GAN) and its applications in classification of mental disorders. J Neurosci Methods. 2020;341:108756.

420. Bi K, Hua L, Wei M, Qin J, Lu Q, Yao Z. Dynamic functional-structural coupling within acute functional state change phases: Evidence from a depression recognition study. J Affect Disord. 2016;191:145–55.

421. Zhu X, Yuan F, Zhou G, Nie J, Wang D, Hu P, Ouyang L, Kong L, Liao W. Cross-network interaction for diagnosis of major depressive disorder based on resting state functional connectivity. Brain Imaging Behav. 2021;15(3):1279–89.

422. Sundermann B, Feder S, Wersching H, Teuber A, Schwindt W, Kugel H, Heindel W, Arolt V, Berger K, Pfleiderer B. Diagnostic classification of unipolar depression based on resting-state functional connectivity MRI: effects of generalization to a diverse sample. J Neural Transm (Vienna, Austria : 1996). 2017;124(5):589–605.

423. Qin J, Wei M, Liu H, Chen J, Yan R, Hua L, Zhao K, Yao Z, Lu Q. Abnormal hubs of white matter networks in the frontal-parieto circuit contribute to depression discrimination via pattern classification. Magn Reson Imaging. 2014;32(10):1314–20.

424. Guo M, Wang T, Zhang Z, Chen N, Li Y, Wang Y, Yao Z, Hu B. Diagnosis of major depressive disorder using whole-brain effective connectivity networks derived from resting-state functional MRI. J Neural Eng. 2020;17(5):056038.

425. Khan DM, Yahya N, Kamel N, Faye I. Automated Diagnosis of Major Depressive Disorder Using Brain Effective Connectivity and 3D Convolutional Neural Network. In: IEEE Access, vol. 9. 2021. p 8835–8846. https://doi.org/10.1109/ACCESS.2021.3049427.

426. Zhang B, Yan G, Yang Z, Su Y, Wang J, Lei T. Brain functional networks based on resting-state EEG data for major depressive disorder analysis and classification. IEEE Trans Neural Syst Rehabil Eng. 2021;29:215–29.

427. Zhu Y, Jayagopal JK, Mehta RK, Erraguntla M, Nuamah J, McDonald AD, Taylor H, Chang SH. Classifying major depressive disorder using fNIRS during motor rehabilitation. IEEE Trans Neural Syst Rehabil Eng. 2020;28(4):961–9.

428. Geng X, Xu J, Liu B, Shi Y. Multivariate classification of major depressive disorder using the effective connectivity and functional connectivity. Front Neurosci. 2018;12:38.

429. Sacchet MD, Prasad G, Foland-Ross LC, Thompson PM, Gotlib IH. Elucidating brain connectivity networks in major depressive disorder

using classification-based scoring. Proc IEEE Int Symp Biomed Imaging. 2014;2014:246–9.

430. Yan B, Xu X, Liu M, Zheng K, Liu J, Li J, Wei L, Zhang B, Lu H, Li B. Quantitative identification of major depression based on resting-state dynamic functional connectivity: a machine learning approach. Front Neurosci. 2020;14:191.

431. Nakano T, Takamura M, Ichikawa N, Okada G, Okamoto Y, Yamada M, Suhara T, Yamawaki S, Yoshimoto J. Enhancing multi-center generalization of machine learning-based depression diagnosis from resting-state fMRI. Front Psych. 2020;11:400.

432. Sacchet MD, Prasad G, Foland-Ross LC, Thompson PM, Gotlib IH. Support vector machine classification of major depressive disorder using diffusion-weighted neuroimaging and graph theory. Front Psych. 2015;6:21.

433. Shi Y, Zhang L, He C, Yin Y, Song R, Chen S, Fan D, Zhou D, Yuan Y, Xie C, et al. Sleep disturbance-related neuroimaging features as potential biomarkers for the diagnosis of major depressive disorder: a multicenter study based on machine learning. J Affect Disord. 2021;295:148–55.

434. Duan L, Duan H, Qiao Y, Sha S, Qi S, Zhang X, Huang J, Huang X, Wang C. Machine learning approaches for MDD detection and emotion decoding using EEG signals. Front Hum Neurosci. 2020;14:284.

435. Saeedi A, Saeedi M, Maghsoudi A, Shalbaf A. Major depressive disorder diagnosis based on effective connectivity in EEG signals: a convolutional neural network and long short-term memory approach. Cogn Neurodyn. 2021;15(2):239–52.

436. Bi K, Chattun MR, Liu X, Wang Q, Tian S, Zhang S, Lu Q, Yao Z. Abnormal early dynamic individual patterns of functional networks in low gamma band for depression recognition. J Affect Disord. 2018;238:366–74.

437. Qin J, Wei M, Liu H, Chen J, Yan R, Yao Z, Lu Q. Altered anatomical patterns of depression in relation to antidepressant treatment: evidence from a pattern recognition analysis on the topological organization of brain networks. J Affect Disord. 2015;180:129–37.

438. Knott V, Mahoney C, Kennedy S, Evans K. EEG power, frequency, asymmetry and coherence in male depression. Psychiatry Res. 2001;106(2):123–40.

439. Marquand AF, Mourão-Miranda J, Brammer MJ, Cleare AJ, Fu CH. Neuroanatomy of verbal working memory as a diagnostic biomarker for depression. NeuroReport. 2008;19(15):1507–11.

440. Lu Q, Bi K, Liu C, Luo G, Tang H, Yao Z. Predicting depression based on dynamic regional connectivity: a windowed Granger causality analysis of MEG recordings. Brain Res. 2013;1535:52–60.

441. Craddock RC, Holtzheimer PE 3rd, Hu XP, Mayberg HS. Disease state prediction from resting state functional connectivity. Magn Reson Med. 2009;62(6):1619–28.

442. Kang M, Kwon H, Park JH, Kang S, Lee Y. Deep-asymmetry: asymmetry matrix image for deep learning method in pre-screening depression. Sensors (Basel). 2020;20(22):6526.

443. Lois G, Wessa M. Differential association of default mode network connectivity and rumination in healthy individuals and remitted MDD patients. Soc Cogn Affect Neurosci. 2016;11(11):1792–801.

444. Hasanzadeh F, Mohebbi M, Rostami R. Graph theory analysis of directed functional brain networks in major depressive disorder based on EEG signal. J Neural Eng. 2020;17(2):026010.

445. Mumtaz W, Ali SSA, Yasin MAM, Malik AS. A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder (MDD). Med Biol Eng Comput. 2018;56(2):233–46.

446. Frick A, Gingnell M, Marquand AF, Howner K, Fischer H, Kristiansson M, Williams SC, Fredrikson M, Furmark T. Classifying social anxiety disorder using multivoxel pattern analyses of brain function and structure. Behav Brain Res. 2014;259:330–5.

447. Zhang W, Yang X, Lui S, Meng Y, Yao L, Xiao Y, Deng W, Zhang W, Gong Q. Diagnostic prediction for social anxiety disorder via multivariate pattern analysis of the regional homogeneity. Biomed Res Int. 2015;2015:763965.

448. Liu F, Guo W, Fouche JP, Wang Y, Wang W, Ding J, Zeng L, Qiu C, Gong Q, Zhang W, et al. Multivariate classification of social anxiety disorder using whole brain functional connectivity. Brain Struct Funct. 2015;220(1):101–15.

449. Xing M, Fitzgerald JM, Klumpp H. Classification of social anxiety disorder with support vector machine analysis using neural correlates of social signals of threat. Front Psych. 2020;11:144.

450. Gavrilescu M, Vizireanu N. Predicting depression, anxiety, and stress levels from videos using the facial action coding system. Sensors (Basel). 2019;19(17):3693.

451. Xie Y, Yang B, Lu X, Zheng M, Fan C, Bi X, Zhou S, Li Y. Anxiety and depression diagnosis method based on brain networks and convolutional neural networks. Annu Int Conf IEEE Eng Med Biol Soc. 2020;2020:1503–6.

452. Qiao J, Li A, Cao C, Wang Z, Sun J, Xu G. Aberrant functional network connectivity as a biomarker of generalized anxiety disorder. Front Hum Neurosci. 2017;11:626.

453. Xing X, Jin L, Li Q, Yang Q, Han H, Xu C, Wei Z, Zhan Y, Zhou XS, Xue Z, et al. Modeling essential connections in obsessive-compulsive disorder patients using functional MRI. Brain Behav. 2020;10(2):e01499.

454. Yang X, Hu X, Tang W, Li B, Yang Y, Gong Q, Huang X. Multivariate classification of drug-naive obsessive-compulsive disorder patients and healthy controls by applying an SVM to resting-state functional MRI data. BMC Psychiatry. 2019;19(1):210.

455. Bruin WB, Taylor L, Thomas RM, Shock JP, Zhutovsky P, Abe Y, Alonso P, Ameis SH, Anticevic A, Arnold PD, et al. Structural neuroimaging biomarkers for obsessive-compulsive disorder in the ENIGMA-OCD consortium: medication matters. Transl Psychiatry. 2020;10(1):342.

456. Takagi Y, Sakai Y, Lisi G, Yahata N, Abe Y, Nishida S, Nakamae T, Morimoto J, Kawato M, Narumoto J, et al. A neural marker of obsessive-compulsive disorder from whole-brain functional connectivity. Sci Rep. 2017;7(1):7538.

457. Zhou C, Cheng Y, Ping L, Xu J, Shen Z, Jiang L, Shi L, Yang S, Lu Y, Xu X. Support vector machine classification of obsessive-compulsive disorder based on whole-brain volumetry and diffusion tensor imaging. Front Psych. 2018;9:524.

458. Hu X, Liu Q, Li B, Tang W, Sun H, Li F, Yang Y, Gong Q, Huang X. Multivariate pattern analysis of obsessive-compulsive disorder using structural neuroanatomy. Eur Neuropsychopharmacol. 2016;26(2):246–54.

459. Bu X, Hu X, Zhang L, Li B, Zhou M, Lu L, Hu X, Li H, Yang Y, Tang W, et al. Investigating the predictive value of different resting-state functional MRI parameters in obsessive-compulsive disorder. Transl Psychiatry. 2019;9(1):17.

460. Sen B, Bernstein GA, Tingting X, Mueller BA, Schreiner MW, Cullen KR, Parhi KK. Classification of obsessive-compulsive disorder from resting-state fMRI. Annu Int Conf IEEE Eng Med Biol Soc. 2016;2016:3606–9.

461. Trambaiolli LR, Biazoli CE Jr, Balardin JB, Hoexter MQ, Sato JR. The relevance of feature selection methods to the classification of obsessive-compulsive disorder based on volumetric measures. J Affect Disord. 2017;222:49–56.

462. Li F, Huang X, Tang W, Yang Y, Li B, Kemp GJ, Mechelli A, Gong Q. Multivariate pattern analysis of DTI reveals differential white matter in individuals with obsessive-compulsive disorder. Hum Brain Mapp. 2014;35(6):2643–51.

463. Shenas SK, Halici U, Çiçek M. A comparative analysis of functional connectivity data in resting and task-related conditions of the brain for disease signature of OCD. Annu Int Conf IEEE Eng Med Biol Soc. 2014;2014:978–81.

464. Liu J, Bu X, Hu X, Li H, Cao L, Gao Y, Liang K, Zhang L, Lu L, Hu X, et al. Temporal variability of regional intrinsic neural activity in drug-naïve patients with obsessive-compulsive disorder. Hum Brain Mapp. 2021;42(12):3792–803.

465. Liu W, Hua M, Qin J, Tang Q, Han Y, Tian H, Lian D, Zhang Z, Wang W, Wang C, et al. Disrupted pathways from frontal-parietal cortex to basal ganglia and cerebellum in patients with unmedicated obsessive compulsive disorder as observed by whole-brain resting-state effective connectivity analysis - a small sample pilot study. Brain Imaging Behav. 2021;15(3):1344–54.

466. Chen Y, Ou Y, Lv D, Ma J, Zhan C, Yang R, Jia C, Shang T, Sun L, Wang Y, et al. Decreased nucleus accumbens connectivity at rest in medication-free patients with obsessive-compulsive disorder. Neural Plast. 2021;2021:9966378.

467. Kalatzis I, Piliouras N, Glotsos D, Ventouras E, Papageorgiou C, Rabavilas A, Soldatos C, Cavouras D. Identifying differences in the P600 component of ERP-signals between OCD patients and controls employing a

PNN-based majority vote classification scheme. Conf Proc IEEE Eng Med Biol Soc. 2005;2005:3994–7.

468. Aydin S, Arica N, Ergul E, Tan O. Classification of obsessive compulsive disorder by EEG complexity and hemispheric dependency measurements. Int J Neural Syst. 2015;25(3):1550010.

469. Zilcha-Mano S, Zhu X, Suarez-Jimenez B, Pickover A, Tal S, Such S, Marohasy C, Chrisanthopoulos M, Salzman C, Lazarov A, et al. Diagnostic and predictive neuroimaging biomarkers for posttraumatic stress disorder. Biol Psychiatry Cogn Neurosci Neuroimaging. 2020;5(7):688–96.

470. Nicholson AA, Densmore M, McKinnon MC, Neufeld RWJ, Frewen PA, Théberge J, Jetly R, Richardson JD, Lanius RA. Machine learning multivariate pattern analysis predicts classification of posttraumatic stress disorder and its dissociative subtype: a multimodal neuroimaging approach. Psychol Med. 2019;49(12):2049–59.

471. Zhu H, Yuan M, Qiu C, Ren Z, Li Y, Wang J, Huang X, Lui S, Gong Q, Zhang W, et al. Multivariate classification of earthquake survivors with post-traumatic stress disorder based on large-scale brain networks. Acta Psychiatr Scand. 2020;141(3):285–98.

472. Shim M, Jin MJ, Im CH, Lee SH. Machine-learning-based classification between post-traumatic stress disorder and major depressive disorder using P300 features. Neuroimage Clin. 2019;24:102001.

473. Nicholson AA, Harricharan S, Densmore M, Neufeld RWJ, Ros T, McKinnon MC, Frewen PA, Théberge J, Jetly R, Pedlar D, et al. Classifying heterogeneous presentations of PTSD via the default mode, central executive, and salience networks with machine learning. NeuroImage Clin. 2020;27:102262.

474. Harricharan S, Nicholson AA, Thome J, Densmore M, McKinnon MC, Théberge J, Frewen PA, Neufeld RWJ, Lanius RA. PTSD and its dissociative subtype through the lens of the insula: Anterior and posterior insula resting-state functional connectivity and its predictive validity using machine learning. Psychophysiology. 2020;57(1):e13472.

475. Park SM, Jeong B, Oh DY, Choi CH, Jung HY. Identification of major psychiatric disorders from resting-state electroencephalography using a machine learning approach. Front Psych. 2021;12:707581.

476. Eken A, Çolak B, Bal NB, Kuşman A, Kızılpınar S, Akaslan DS, Baskak B. Hyperparameter-tuned prediction of somatic symptom disorder using functional near-infrared spectroscopy-based dynamic functional connectivity. J Neural Eng. 2019;17(1):016012.

477. Lavagnino L, Amianto F, Mwangi B, D'Agata F, Spalatro A, Zunta-Soares GB, Abbate Daga G, Mortara P, Fassino S, Soares JC. Identifying neuroanatomical signatures of anorexia nervosa: a multivariate machine learning approach. Psychol Med. 2015;45(13):2805–12.

478. Lavagnino L, Mwangi B, Cao B, Shott ME, Soares JC, Frank GKW. Cortical thickness patterns as state biomarker of anorexia nervosa. Int J Eat Disord. 2018;51(3):241–9.

479. Geisler D, Borchardt V, Boehm I, King JA, Tam FI, Marxen M, Biemann R, Roessner V, Walter M, Ehrlich S. Altered global brain network topology as a trait marker in patients with anorexia nervosa. Psychol Med. 2020;50(1):107–15.

480. Weygandt M, Schaefer A, Schienle A, Haynes JD. Diagnosing different binge-eating disorders based on reward-related brain activation patterns. Hum Brain Mapp. 2012;33(9):2135–46.

481. Lee MH, Kim N, Yoo J, Kim HK, Son YD, Kim YB, Oh SM, Kim S, Lee H, Jeon JE, et al. Multitask fMRI and machine learning approach improve prediction of differential brain activity pattern in patients with insomnia disorder. Sci Rep. 2021;11(1):9402.

482. Jansen C, Penzel T, Hodel S, Breuer S, Spott M, Krefting D. Network physiology in insomnia patients: Assessment of relevant changes in network topology with interpretable machine learning models. Chaos (Woodbury, NY). 2019;29(12):123129.

483. Zhang J, Liu Y, Luo R, Du Z, Lu F, Yuan Z, Zhou J, Li S. Classification of pure conduct disorder from healthy controls based on indices of brain networks during resting state. Med Biol Eng Comput. 2020;58(9):2071–82.

484. Zhang J, Cao W, Wang M, Wang N, Yao S, Huang B. Multivoxel pattern analysis of structural MRI in children and adolescents with conduct disorder. Brain Imaging Behav. 2019;13(5):1273–80.

485. Zhang J, Liu W, Zhang J, Wu Q, Gao Y, Jiang Y, Gao J, Yao S, Huang B. Distinguishing adolescents with conduct disorder from typically

developing youngsters based on pattern classification of brain structural MRI. Front Hum Neurosci. 2018;12:152.

486. Tang Y, Jiang W, Liao J, Wang W, Luo A. Identifying individuals with antisocial personality disorder using resting-state FMRI. PLoS One. 2013;8(4):e60652.

487. Tang Y, Liu W, Chen J, Liao J, Hu D, Wang W. Altered spontaneous activity in antisocial personality disorder revealed by regional homogeneity. NeuroReport. 2013;24(11):590–5.

488. Sato JR, de Oliveira-Souza R, Thomaz CE, Basílio R, Bramati IE, Amaro E Jr, Tovar-Moll F, Hare RD, Moll J. Identification of psychopathic individuals using pattern classification of MRI images. Soc Neurosci. 2011;6(5–6):627–39.

489. Wetherill RR, Rao H, Hager N, Wang J, Franklin TR, Fan Y. Classifying and characterizing nicotine use disorder with high accuracy using machine learning and resting-state fMRI. Addict Biol. 2019;24(4):811–21.

490. Li Y, Cui Z, Liao Q, Dong H, Zhang J, Shen W, Zhou W. Support vector machine-based multivariate pattern classification of methamphetamine dependence using arterial spin labeling. Addict Biol. 2019;24(6):1254–62.

491. Ding X, Li Y, Li D, Li L, Liu X. Using machine-learning approach to distinguish patients with methamphetamine dependence from healthy subjects in a virtual reality environment. Brain and behavior. 2020;10(11):e01814.

492. Mete M, Sakoglu U, Spence JS, Devous MD Sr, Harris TS, Adinoff B. Successful classification of cocaine dependence using brain imaging: a generalizable machine learning approach. BMC Bioinformatics. 2016;17(Suppl 13):357.

493. Adeli E, Zahr NM, Pfefferbaum A, Sullivan EV, Pohl KM. Novel machine learning identifies brain patterns distinguishing diagnostic membership of human immunodeficiency virus, alcoholism, and their comorbidity of individuals. Biol Psychiatry Cogn Neurosci Neuroimaging. 2019;4(6):589–99.

494. Guggenmos M, Schmack K, Veer IM, Lett T, Sekutowicz M, Sebold M, Garbusow M, Sommer C, Wittchen HU, Zimmermann US, et al. A multimodal neuroimaging classifier for alcohol dependence. Sci Rep. 2020;10(1):298.

495. Mehla VK, Singhal A, Singh P. A novel approach for automated alcoholism detection using Fourier decomposition method. J Neurosci Methods. 2020;346:108945.

496. Zhu X, Du X, Kerich M, Lohoff FW, Momenan R. Random forest based classification of alcohol dependence patients and healthy controls using resting state MRI. Neurosci Lett. 2018;676:27–33.

497. Mumtaz W, Saad M, Kamel N, Ali SSA, Malik AS. An EEG-based functional connectivity measure for automatic detection of alcohol use disorder. Artif Intell Med. 2018;84:79–89.

498. Wang SH, Lv YD, Sui Y, Liu S, Wang SJ, Zhang YD. Alcoholism detection by data augmentation and convolutional neural network with stochastic pooling. J Med Syst. 2017;42(1):2.

499. Kinreich S, McCutcheon VV, Aliev F, Meyers JL, Kamarajan C, Pandey AK, Chorlian DB, Zhang J, Kuang W, Pandey G, et al. Predicting alcohol use disorder remission: a longitudinal multimodal multi-featured machine learning approach. Transl Psychiatry. 2021;11(1):166.

500. Bae Y, Yoo BW, Lee JC, Kim HC. Automated network analysis to measure brain effective connectivity estimated from EEG data of patients with alcoholism. Physiol Meas. 2017;38(5):759–73.

501. Kumar S, Ghosh S, Tetarway S, Sinha RK. Support vector machine and fuzzy C-mean clustering-based comparative evaluation of changes in motor cortex electroencephalogram under chronic alcoholism. Med Biol Eng Comput. 2015;53(7):609–22.

502. Khan DM, Yahya N, Kamel N, Faye I. Effective connectivity in default mode network for alcoholism diagnosis. IEEE Trans Neural Syst Rehabil Eng. 2021;29:796–808.

503. Mumtaz W, Vuong PL, Xia L, Malik AS, Rashid RBA. An EEG-based machine learning method to screen alcohol use disorder. Cogn Neurodyn. 2017;11(2):161–71.

504. Hahn S, Mackey S, Cousijn J, Foxe JJ, Heinz A, Hester R, Hutchinson K, Kiefer F, Korucuoglu O, Lett T, et al. Predicting alcohol dependence from multi-site brain structural measures. Hum Brain Mapp. 2022;43(1):555–65.

505. Zhang H, Silva FHS, Ohata EF, Medeiros AG, Rebouças Filho PP. Bi-dimensional approach based on transfer learning for alcoholism

Chen *et al. BMC Medicine*   (2023) 21:241

Page 29 of 29

pre-disposition classification via EEG signals. Front Hum Neurosci. 2020;14:365.

506. Wang SH, Xie S, Chen X, Guttery DS, Tang C, Sun J, Zhang YD. Alcoholism identification based on an AlexNet transfer learning model. Front Psych. 2019;10:205.

507. Prabhakar SK, Rajaguru H. Alcoholic EEG signal classification with Correlation Dimension based distance metrics approach and Modified Adaboost classification. Heliyon. 2020;6(12):e05689.

508. Erguzel TT, Noyan CO, Eryilmaz G, Ünsalver B, Cebi M, Tas C, Dilbaz N, Tarhan N. Binomial logistic regression and artificial neural network methods to classify opioid-dependent subjects and control group using quantitative EEG power measures. Clin EEG Neurosci. 2019;50(5):303–10.

509. Cremers H, van Zutphen L, Duken S, Domes G, Sprenger A, Waldorp L, Arntz A. Borderline personality disorder classification based on brain network measures during emotion regulation. Eur Arch Psychiatry Clin Neurosci. 2021;271(6):1169–78.

510. Xu T, Cullen KR, Houri A, Lim KO, Schulz SC, Parhi KK. Classification of borderline personality disorder based on spectral power of resting-state fMRI. Annu Int Conf IEEE Eng Med Biol Soc. 2014;2014:5036–9.

511. Chekroud AM, Bondar J, Delgadillo J, Doherty G, Wasil A, Fokkema M, Cohen Z, Belgrave D, DeRubeis R, Iniesta R, et al. The promise of machine learning in predicting treatment outcomes in psychiatry. World Psychiatry. 2021;20(2):154–70.

512. Greene AS, Shen X, Noble S, Horien C, Hahn CA, Arora J, Tokoglu F, Spann MN, Carrión CI, Barron DS, et al. Brain-phenotype models fail for individuals who defy sample stereotypes. Nature. 2022;609(7925):109–18.

513. Li J, Bzdok D, Chen J, Tam A, Ooi LQR, Holmes AJ, Ge T, Patil KR, Jabbi M, Eickhoff SB, et al. Cross-ethnicity/race generalization failure of behavioral prediction from resting-state functional connectivity. Sci Adv. 2022;8(11):eabj1812.

514. Grzenda A, Kraguljac NV, McDonald WM, Nemeroff C, Torous J, Alpert JE, Rodriguez CI, Widge AS. Evaluating the machine learning literature: a primer and user's guide for psychiatrists. Am J Psychiatry. 2021;178(8):715–29.

515. Zadrozny B. Learning and evaluating classifiers under sample selection bias. In: Proceedings of the twenty-first international conference on Machine learning. Banff, Alberta, Canada: Association for Computing Machinery; 2004. p. 114.

516. Marek S, Tervo-Clemmens B, Calabro FJ, Montez DF, Kay BP, Hatoum AS, Donohue MR, Foran W, Miller RL, Hendrickson TJ, et al. Reproducible brain-wide association studies require thousands of individuals. Nature. 2022;603(7902):654–60.

517. Goldfarb MG, Brown DR. Diversifying participation: the rarity of reporting racial demographics in neuroimaging research. Neuroimage. 2022;254:119122.

518. Webb EK, Etter JA, Kwasa JA. Addressing racial and phenotypic bias in human neuroscience methods. Nat Neurosci. 2022;25(4):410–4.

519. Futoma J, Simons M, Panch T, Doshi-Velez F, Celi LA. The myth of generalisability in clinical research and machine learning in health care. Lancet Digital health. 2020;2(9):e489–92.

520. Celi LA, Cellini J, Charpignon ML, Dee EC, Dernoncourt F, Eber R, Mitchell WG, Moukheiber L, Schirmer J, Situ J, et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities-A global review. PLOS Digit Health. 2022;1(3):e0000022.

521. Rutledge RB, Chekroud AM, Huys QJ. Machine learning and big data in psychiatry: toward clinical applications. Curr Opin Neurobiol. 2019;55:152–9.

522. Santos MS, Soares JP, Abreu PH, Araujo H, Santos J. Cross-validation for imbalanced datasets: avoiding overoptimistic and overfitting approaches [Research Frontier]. IEEE Comput Intell Mag. 2018;13(4):59–76.

523. Janssen RJ, Mourão-Miranda J, Schnack HG. Making individual prognoses in psychiatry using neuroimaging and machine learning. Biol Psychiatry Cogn Neurosci Neuroimaging. 2018;3(9):798–808.

524. Ho SY, Phua K, Wong L, Bin Goh WW. Extensions of the external validation for checking learned model interpretability and generalizability. Patterns (New York, NY). 2020;1(8):100129.

525. Cai XL, Xie DJ, Madsen KH, Wang YM, Bögemann SA, Cheung EFC, Møller A, Chan RCK. Generalizability of machine learning for

classification of schizophrenia based on resting-state functional MRI data. Hum Brain Mapp. 2020;41(1):172–84.

526. Dhamala E, Yeo BTT, Holmes AJ. One Size Does Not Fit All: Methodological Considerations for Brain-Based Predictive Modeling in Psychiatry. Biol Psychiatry. 2023;93(8):717–28. https://doi.org/10.1016/j.biopsych.2022.09.024.

527. Dinsdale NK, Bluemke E, Sundaresan V, Jenkinson M, Smith SM, Namburete AIL: Challenges for machine learning in clinical translation of big data imaging studies. Neuron.

528. Leenings R, Winter NR, Dannlowski U, Hahn T. Recommendations for machine learning benchmarks in neuroimaging. Neuroimage. 2022;257:119298.

529. Fusar-Poli P, Hijazi Z, Stahl D, Steyerberg EW. The science of prognosis in psychiatry: a review. JAMA Psychiat. 2018;75(12):1289–97.

Publisher's Note

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