PCN Psychiatry and Clinical Neurosciences

REVIEW ARTICLE

Unveiling the potential of machine learning in schizophrenia diagnosis: A meta-analytic study of task-based neuroimaging data

Xuan Wang, MS,[1,2,3,4†] Chao Yan, PhD ,[1,2†,*] Peng-yuan Yang, MS,[5] Zheng Xia, BS,[1] Xin-lu Cai, PhD,[6] Yi Wang, PhD,[3,4] Sze Chai Kwok, DPhil [1,2,7,8] and Raymond C.K. Chan, PhD [3,4]

The emergence of machine learning (ML) techniques has opened up new avenues for identifying biomarkers associated with schizophrenia (SCZ) using task-related fMRI (t-fMRI) designs. To evaluate the effectiveness of this approach, we conducted a comprehensive meta-analysis of 31 t-fMRI studies using a bivariate model. Our findings revealed a high overall sensitivity of 0.83 and specificity of 0.82 for t-fMRI studies. Notably, neuropsychological domains modulated the classification performance, with selective attention demonstrating a significantly higher specificity than working memory ($\beta = 0.98$, $z = 2.11$, $P = 0.04$). Studies involving older, chronic patients with SCZ reported higher sensitivity ($P < 0.015$) and specificity ($P < 0.001$) than those involving younger, first-episode patients or high-risk individuals for psychosis. Additionally, we found that the severity of negative symptoms was positively associated with the specificity of the classification model ($\beta = 7.19$, $z = 2.20$, $P = 0.03$). Taken together, these results support the potential of using task-based fMRI data in combination with machine learning techniques to identify biomarkers related to symptom outcomes in SCZ, providing a promising avenue for improving diagnostic accuracy and treatment efficacy. Future attempts to deploy ML classification should consider the factors of algorithm choice, data quality and quantity, as well as issues related to generalization.

Keywords: *attention, machine learning, meta-analysis, schizophrenia, task-based fMRI.*

http://onlinelibrary.wiley.com/doi/10.1111/pcn.13625/full

Schizophrenia (SCZ) is a complex and heterogeneous mental disorder, such that researchers have encountered great challenges in their quest for finding sensitive and reliable biomarkers to improve diagnostic accuracy.[1] Over the past few decades, neuroscientists and psychiatrists have explored various methods to identify biomarkers associated with SCZ symptoms. The emergence of machine learning (ML) has offered a promising avenue for detecting specific brain patterns related to SCZ symptoms, thereby potentially revolutionizing the current diagnostic process.[2–4]

Machine learning (ML) algorithms, such as support vector machine (SVM), logistic regression (LGR), and linear discriminant analysis (LDA), have been applied to distinguish patients with schizophrenia (SCZ)[5,6] from healthy controls (HC), as well as those at risk of developing SCZ, such as individuals with schizotypy traits or genetic risks.[7,8] These algorithms take into account the multivariate nature of changes in brain structure and function, as measured by diverse neuroimaging modalities including structural magnetic resonance imaging (sMRI) (e.g. T1-weighted imaging),[9,10] diffusion tensor imaging (DTI),[11–13] and functional magnetic resonance imaging (fMRI).[14,15] For instance, a meta-analysis study by Kambeitz *et al*. (2015) revealed that resting-state functional magnetic resonance imaging (r-fMRI) and structural MRI (sMRI) studies achieved high levels of sensitivity and specificity in distinguishing SCZ patients from HC.[3] However, the potential of task-related fMRI (t-fMRI) modalities and their underlying influential factors for achieving better diagnostic outcomes remains largely unclear. Interestingly, evidence suggests that t-fMRI may provide a more comprehensive characterization of the psychotic symptoms and behavioral deficits seen in SCZ.[16–18] As such, exploring the use of t-fMRI covering multiple facets of cognitive, emotional, psychosocial processing in ML algorithms could yield novel insights into the neurobiological mechanisms of SCZ and may improve diagnostic accuracy.[3,7] This meta-analysis aimed to provide a comprehensive assessment of the effectiveness of machine learning classification based on t-fMRI data in the diagnosis of patients with SCZ.

Notably, several factors, including the neuropsychological domains, demographic characteristics, and clinical features of participants, may influence the classification performance of t-fMRI studies.[19] For instance, previous machine learning studies focused on

[1] Key Laboratory of Brain Functional Genomics (MOE&STCSM), Affiliated Mental Health Center (ECNU), School of Psychology and Cognitive Science, East China Normal University, Shanghai, China
[2] Shanghai Changning Mental Health Center, Shanghai, China
[3] Neuropsychology and Applied Cognitive Neuroscience Laboratory; CAS Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China
[4] Department of Psychology, University of Chinese Academy of Sciences, Beijing, China
[5] Faculty of Science, Ghent University, Ghent, Belgium
[6] Institute of Brain Science and Department of Physiology, School of Basic Medical Sciences, Hangzhou Normal University, Hangzhou, China
[7] Phylo-Cognition Laboratory, Division of Natural and Applied Sciences, Data Science Research Center, Duke Kunshan University, Kunshan, China
[8] Shanghai Key Laboratory of Magnetic Resonance, East China Normal University, Shanghai, China
* Correspondence: Email: cyan@psy.ecnu.edu.cn
† These authors contributed equally to this work.

various neuropsychological domains, such as selective attention, working memory, and emotional recognition, have reported large variations in the performance of the ML algorithms. The study conducted by Demirci *et al*. examined the effectiveness of machine learning (ML) algorithms in three different neuropsychological domains.[20] Their findings demonstrated that the level of sensitivity varied across domains, with the attention domain achieving a sensitivity of 0.79, the memory domain achieving 0.77, and the sensorimotor interaction domain achieving 0.95. Additionally, different sub-components of the same neuropsychological domain may have different classification performances. For instance, two separate studies[21,22] employed verbal and spatial N-back tasks to classify patients with SCZ. However, classification performance varied significantly across the two studies, with the accuracy ranging from the lowest of .50[21] to the highest of .94.[22]

Additionally, research has shown that classification performance in t-fMRI studies may be influenced by a range of demographic and clinical factors, including age, sex, illness stage, medication dosage, and the presence of psychotic symptoms.[3] Furthermore, classification outcomes can also be influenced by various ML-related issues, such as the choice of classification algorithms, the selection of features, sample size, and the use of multiple sites.[23–25] Thus, it is crucial to take these factors into account to better understand their potential influence on t-fMRI classification performance.

In this study, we sought to examine the overall classification performance of the ML algorithms for SCZ across the t-fMRI studies and explore the potential moderation effect of neuropsychological domains, demographics, clinical characteristics, and ML approaches on classification performances. To this end, we used a bivariate meta-analysis model[26] to consider both sensitivity and specificity in its classification.

Methods

Search strategy and selection criteria

Adhering to the PRISMA international guidelines,[27] we systematically searched the Web of Science and PubMed databases to identify pertinent studies published from January 1980 to September 2022. The keywords for the comprehensive search included terms of ML algorithms (i.e. "support vector" OR "SVM" OR "classification" OR "categorization" OR "single diagnosis") AND terms of MRI modality (i.e. "fMRI" OR "functional magnetic resonance imaging") AND terms of type of fMRI study (i.e. "task") AND terms of participants (i.e. "SCZ" OR "psychosis" OR "psychotic"). After removing the duplicated studies (K = 54) and irrelevant studies (K = 32), 140 studies were further scrutinized, based on the following exclusion criteria: (i) reviews or meta-analysis articles; (ii) classification studies that used other modalities (e.g. structure MRI or resting-state fMRI) rather than t-fMRI; (iii) classification studies that were not used for diagnosing SCZ; (iv) classification studies that distinguished different cognitive components at the individual level; and (v) classification studies that reported insufficient data for estimating sensitivity and specificity. Two independent assessors accomplished the above screening processes. In addition to the articles initially selected, we incorporated three more articles[16,20,21] from a review[20] and one study[8] that fulfilled the specified criteria. As a result, the final meta-analysis comprised 30 papers (See Fig. 1).

Data coding

For each study, three independent assessors documented the ML algorithm's classification parameters (e.g. TP, FP, TN, FN), sample size, experimental paradigms, demographic and clinical variables (e.g. age, sex ratio, illness stage, medication usage), and severity of positive and negative symptoms (see Table 1). Additionally, we considered various ML factors, such as classification algorithms (e.g. support-vector machine, logistic regression, deep learning, etc.), features (e.g. regional activation *vs*. functional connectivity), feature selection/extraction approach (independent component analysis, principal component analysis, etc.), cross-validation approach, and data source (single site *vs*. multiple sites). We evaluated the classification performance of different ML algorithms for distinguishing SCZ patients from HC using sensitivity (TP/(TP + FN)) and specificity measures (TN/(TN + FP)).

Calculation of summarized effect size

We used the bivariate model proposed by Reitsma[26] to conduct a meta-analysis, assuming that sensitivity and specificity followed a binomial normal distribution.[28] A mixed linear model was applied to estimate the summarized effect values.[26] We assessed heterogeneity using a chi-square test, with the null hypothesis that all performances were equal. To account for two datasets being used repeatedly in four studies,[29–32] we calculated the average effect size for these two datasets.

Publication bias

To evaluate the possibility of publication bias, we employed several methods. First, we visually inspected the funnel plot for any asymmetry. Additionally, we conducted the Deek's test, which is a statistical test that examines the significance of funnel plot asymmetry.[33] Furthermore, we performed Orwin's fail-safe N analyses[34,35] to assess

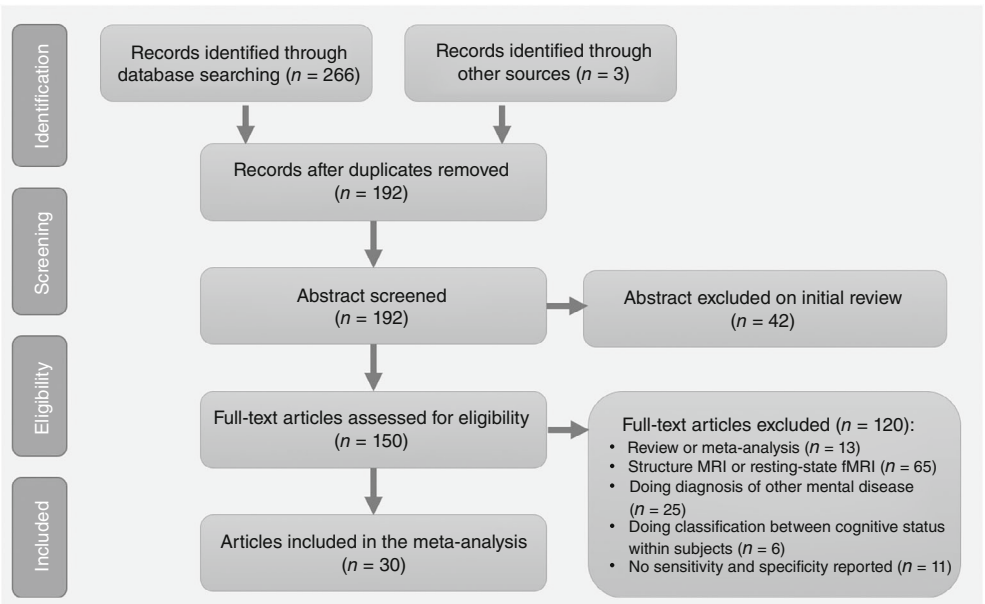

Fig. 1 PRISMA flow diagram of article identification.

Table 1. Descriptive statistics of studies included in the meta-analysis

Studies	Task	Participants	Diagnostic criteria	N (SCZ/HR)	N (HC)	Age (SCZ/HR)	Age (HC)	Male % (SCZ)	Male % (HC)	DOI (years)	Classification algorithm	Cross-validation approach
Demirci *et al.* 2008a[83]	AOD	SCZ	DSM-IV	34	36	31.4	28.9	0.82	0.78	/	PPA	LOOCV
Calhoun *et al.* 2008[16]	AOD	SCZ	DSM-IV	21	40	34.9	31.3	0.71	0.73	12.4	NBA	LOOCV
Demirci *et al.* 2008b[20]	AOD	SCZ	/	57	91	32.4	30.2	0.67	0.66	/	PPA	LOOCV
	SIRP	SCZ	/	57	91	32.4	30.2	0.67	0.66	/	PPA	LOOCV
	SM	SCZ	/	57	91	32.4	30.2	0.67	0.66	/	PPA	LOOCV
Sui *et al.* 2009[84]	AOD + SIRP +SM	SCZ	/	25	28	32	32	0.88	0.82	/	NBA	LOOCV
Yang *et al.*, 2010[85]	AOD	SCZ	DSM-IV	20	20	40.2	42.5	0.85	0.6	/	SVM with Majority voting	LOOCV
Costafreda *et al.* 2011[86]	VFT	SCZ	DSM-IV	32	40	35.5	35.8	0.81	0.5	11.4	Multi-class linear SVM	LOOCV
Du *et al.* 2012[87]	AOD	SCZ	/	28	28	39.4	31.5	0.82	0.68	/	NBA	LOOCV
Modinos *et al.* 2012[7]	EET	PP	CAPE	18	18	19.7	21	0.5	0.5	/	Linear SVM	LOOCV
Yoon *et al.* 2012[73]	AXCPT	SCZ	DSM-IV	51	51	19.9	20.2	0.76	0.51	/	LDA	LOOCV
Rish *et al.* 2013[88]	AAT	SCZ	DSM-IV	11	11	34	34	/	/	12	GNB, **Linear SVM**, RF	LOOCV
Castro *et al.* 2014[89]	AOD	SCZ	DSM-IV	31	21	27.7	26.6	0.68	0.38	/	ν-MKL	LOOCV
Juneja *et al.* 2014[29]	AOD	SCZ	/	34	34	39.8	37.9	/	/	/	Linear SVM	LOOCV
	AOD	SCZ	/	25	25	44.5	40.1	/	/	/	Linear SVM	LOOCV
Mascolo *et al.* 2014[90]	Spatial N-back	SCZ	/	9	9	/	/	/	/	/	Linear SVM	LOOCV
Bendfeldt *et al.*, 2015[21]	Verbal N-back	ARMS	ICD-10	19	19	/	/	/	/	/	Linear SVM	LOOCV
	Verbal N-back	SCZ	ICD-10	19	19	/	/	/	/	/	Linear SVM	LOOCV
Gaebler *et al.* 2015[91]	MMT	SCZ	DSM-IV	24	24	36.1	36.4	0.58	0.58	8.3	Linear SVM	LOOCV
Koch *et al.* 2015[92]	MID	SCZ	/	44	44	34.2	37.1	0.61	0.8	4.8	Linear SVM	LOOCV
Juneja *et al.*, 2016[30]	AOD	SCZ	/	34	34	39.8	37.9	/	/	/	Linear SVM	LOOCV
	AOD	SCZ	/	25	25	44.5	40.1	/	/	/	Linear SVM	LOOCV
Valli *et al.* 2016[8]	VMT	ARMS	PACE	25	25	23.84	25.12	0.72	0.56	/	Linear SVM	LOOCV
Orban *et al.* 2018[5]	EMT + SIRP +N-back +AOD	SCZ	/	191	191	33.4	32.3	0.73	0.75	/	Linear SVM	10-fold CV
Bae *et al.*, 2018[6]	N-back	SCZ	DSM-IV	21	54	24.26	22.51	0.71	0.57	/	NBA, **Linear SVM**, DT, GNB and LDA	10-fold CV
Juneja *et al.*, 2018a[32]	AOD	SCZ	/	34	34	39.8	37.9	/	/	/	Linear SVM	LOOCV
	AOD	SCZ	/	25	25	44.5	40.1	/	/	/	Linear SVM	LOOCV
Juneja *et al.*, 2018b[31]	AOD	SCZ	/	34	34	39.8	37.9	/	/	/	Linear SVM	LOOCV
	AOD	SCZ	/	25	25	44.5	40.1	/	/	/	Linear SVM	LOOCV
Viviano *et al.* 2018[93]	PERT	SSD	/	114	74	34.1	31.3	/	/	/	Linear SVM	10-fold CV
Lombardi *et al.*, 2019[94]	Spatial N-back	SCZ	DSM-IV	42	49	34	30	0.76	0.53	/	Linear SVM	10-fold CV
Oh *et al.*, 2019[95]	IRET	SCZ	DSM-IV	103	41	32.46	33.98	0.55	0.59	4.69	DL, SVM	10-fold CV
Antonucci *et al.*, 2020[96]	VACT	SCZ	DSM-IV	65	65	29	29	0.77	0.82	/	SVM with Majority voting	Reversal discovery-validation strategy in the HC-SCZ and HC-SIB cohort
	VACT	SCZ	/	39	39	35	35	0.38	0.53	-	SVM with Majority voting	Reversal discovery-validation strategy in the HC-SCZ and HC-SIB cohort

Table 1. (*Continued*)

Studies	Task	Participants	Diagnostic criteria	N (SCZ/HR)	N (HC)	Age (SCZ/HR)	Age (HC)	Male % (SCZ)	Male % (HC)	DOI (years)	Classification algorithm	Cross-validation approach
Yang *et al.*, 2020[97]	SST	SCZ	DSM-IV	44	44	36.86	36.73	0.75	0.75	/	Linear SVM	LOOCV
Ghosal *et al.*, 2021[22]	Spatial N-back	SCZ	/	53	53	/	/	/	/	/	**LGR,** Linear SVM, RF	10-fold CV
	Spatial N-back	SCZ	/	43	54	/	/	/	/	/	**LGR,** Linear SVM, RF	10-fold CV
Potvin *et al.*, 2021[72]	MRT	SCZ	DSM-IV	30	30	32.5	30.5	0.51	0.48	9.8 ± 7.1	LGR (LASSO; Ridge regularization), RF, **Linear SVM**, LDA	LOOCV
Cruz-Martinez *et al.*, 2022[98]	AOD	SCZ	/	14	14	32	36.71	/	/	/	SVM with recursive feature eliminating	LOOCV

Note: ML algorithms written in bold reported the best classification performances among all the approaches utilized in the corresponding studies. Abbreviations: 10-fold CV, 10-fold cross validation; AAT, Auditory attention task; AOD, Auditory oddball task; AXCPT, AX-version of the Continuous Performance Task; CAPE, the Community Assessment of Psychic Experiences questionnaire; DL, Deep learning; DOI, Duration of illness; DT, Decision tree; EET, Emotion experience task; EMT, Emotion-memory task; GNB, Gaussian naïve Bayes approach; HC, Healthy controls; IRET, Ideas of reference eliciting task; LDA, Linear Discriminant Analysis; LGR, Logistic regression; LOOCV, Leave-one-out cross validation; LSR, Lasso regression; MID, Monetary incentive delay task; MKL, Multiple Kernel Learning; MMT, Mismatch task; MRT, Mental rotation task; NBA, Neighbor-based algorithm; Num, Number; PACE, Personal Assessment and Crisis Evaluation; PERT, Penn Emotion Recognition task; PPA, Projection pursuit algorithm; RF, Random forests; SCAN, Schedules for clinical assessment in neuropsychiatry; SCZ, Schizophrenia patients; SIRP, Sternberg item recognition paradigm; SM, Sensory motor task; SSD, Schizophrenia Spectrum Disorder; SSD, Schizophrenia Spectrum disorders; SST, Stop signal task; VACT, Variable attention control task; VFT, Verbal fluency task; VMT, Verbal memory task.

the robustness of our findings by estimating the number of studies with null results that would be needed to nullify the observed effect size.

Moderator analyses

To investigate the impact of various factors on the chosen classification performance indicators, we conducted moderator analyses. The factors included: neuropsychological domains (such as selective attention *vs*. working memory), demographic factors (such as age and sex ratio), ML approach-related factors (such as the features selected (i.e. regional activation *vs*. brain connection), classification algorithm used (e.g. SVM, LDA, LGR, Random forest (RF), deep learning (DL), Gaussian Naïve Bayes (GNB), etc.), data source (i.e. single site *vs*. multiple sites), and sample size), as well as clinical factors (such as medication dosage, severity of positive and negative symptoms, and illness stage). To account for the variability in the number of items and score ranges across different clinical scales for assessing the severity of symptoms, we calculated a relative symptom score. Specifically, we divided the positive or negative symptom scores by the maximum possible score for the corresponding scale. Moreover, we calculated converted scores for both positive and negative symptoms in accordance with PANSS by utilizing the recommended conversion tool proposed by Erp.[36] Following the illness stage classification method proposed by Fava and Keller,[37] SCZ patients with a duration of illness (DOI) >2 years were categorized as chronic patients. SCZ patients with a DOI of <6 months or having been specifically labeled as first-episode patients (FEP) by the authors of the articles identified were designated as FEP. Additionally, individuals meeting the criteria for high risk (HR) of psychosis based on clinical, genetic, or psychometric factors were classified as being at high risk for psychosis.[38] Given that individuals at high risk for psychosis exhibited comparable cognitive dysfunction and neural abnormalities to those observed in FEP,[39,40] and the limited number of studies for each of these patient groups, we combined the high-risk and the SCZ groups for the moderation analysis. To test for moderation effects, we compared the default linear mixed model with a regression model that included one of these moderators. We dropped any moderator analysis that involved fewer than five studies due to weak statistical power.

To compare the effectiveness of different neuroimaging modalities, we conducted a study where we compared the classification performance of t-fMRI with r-fMRI and sMRI. To achieve this, we analyzed the classification performance data from previous studies on r-fMRI and sMRI, which were reported in a meta-analysis study by Kambeitz *et al.*[3] and compared them with our data on t-fMRI using bivariate moderator analysis. The sMRI studies included 20 studies, with 1040 healthy controls and 970 SCZ patients, while the r-fMRI studies included 11 studies, with 284 healthy controls and 308 patients with SCZ.

All the analyses based on the bivariate model were performed using the package ***Mada*** implemented with the R-language.[41] The Comprehensive Meta-Analysis (CMA) 2.0 software[42] was applied to determine the level of publication bias.

Results

Descriptive statistics

The meta-analysis identified 31 eligible articles, and pooled data from 31 independent studies with sample sizes ranging from 18 to 382, accruing a total of 1286 patients with SCZ (69% male, mean age = 32.66 ± 5.90) and 1291 healthy controls (62% male, mean age = 31.87 ± 5.61). No group difference was observed in age (t (50) = 0.49, P = 0.62), but there was a significant over-representation of males in the SCZ group compared to the healthy control group (χ^2 (1) = 13.68, P < 0.001) (see Table 1).

Summarized effect size

Among these studies (K = 31), t-fMRI-based ML classification algorithm separated patients with SCZ from HC with a sensitivity of 0.83 (95% confidence interval (CI): 0.77–0.87) and a specificity of 0.82

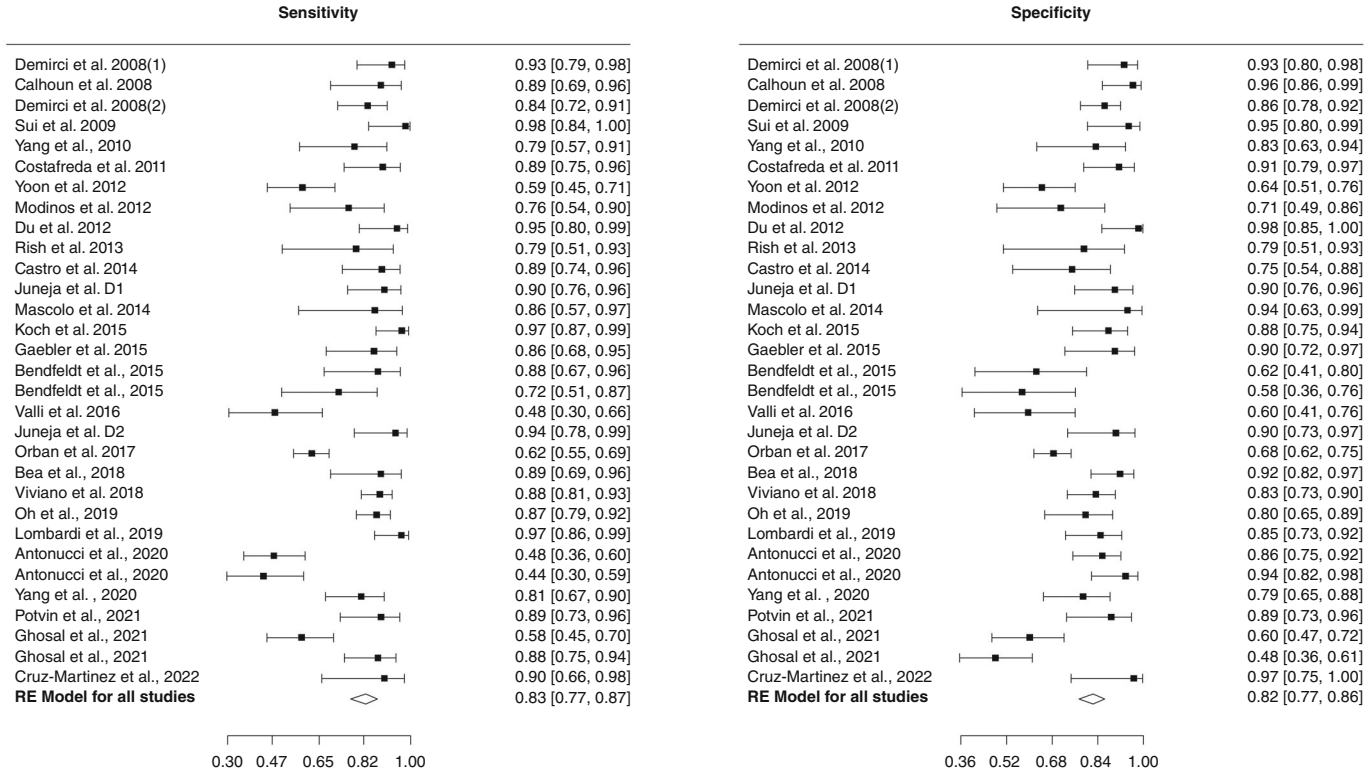

Fig. 2 Forest plots of sensitivity and specificity for each study and summarized estimations calculated by the bivariate model.

(95% CI: 0.77–0.86) with an area under the curve (AUC) of 0.89 (Fig. 2). The large summarized Log Odds Ratio of 3.11 (>2.18) indicated a significant effect size for the classification performance. Tests for equality showed significant differences in both sensitivity (χ^2 (30) = 200.19, $P < 0.001$) and specificity (χ^2 (28) = 138.19, $P < 0.001$), indicating significant heterogeneity in classification performance across the studies.

Publication bias

The funnel plot revealed an asymmetrical pattern, suggesting the possible presence of unpublished small studies and publication bias (see Fig. S1). To further test the degree of asymmetry, we conducted Deek's funnel plot test, which showed a significant result (t (29) = 2.66, $P = 0.013$), indicating that publication bias was likely to be present. The Fail-safe N test indicated that 141 studies with weak effect sizes (Log Odds Ratio <0.18) were needed to nullify the observed effect size (Log Odds Ratio <0.36) (Fig. S1).

Moderation effect of neuropsychological domain and MRI modality

Neuropsychological domain

Fifteen t-fMRI studies were conducted using tasks that targeted at the attention domain, including selective attention and attention/cognitive control. In addition, nine studies explored the memory domain, which included working memory, short-term and long-term memory. Four studies focused on the emotion/social cognition domain, while two studies assessed the sensory-motor/perception domain. Furthermore, one study evaluated the language domain. Two other studies employed multiple tasks that covered various neuropsychological domains (See Table S1).

When analyzing studies that focused on selective attention (K = 11) separately, the meta-analysis yielded a sensitivity of 0.87 (95% CI: 0.82–0.91), specificity of 0.88 (95% CI: 0.83–0.92), and AUC of 0.94, indicating high accuracy in classification performance. Moreover, the equality assumption was satisfied across these studies (Sensitivity: χ^2 (10) = 10.30, $P = 0.41$; Specificity: χ^2 (10) = 15.43, $P = 0.12$). The studies investigating working memory (K = 7) resulted in a sensitivity of 0.84 (95%CI: 0.71–0.91), specificity of 0.72 (95%CI: 0.57–0.84), and AUC of 0.85. However, the between-study variance may be attributed to heterogeneity (Sensitivity: χ^2 (6) = 32.89, $P < 0.001$; Specificity: χ^2 (6) = 36.85, $P < 0.001$). Overall, the bivariate meta-analytic model exhibited superior classification performance in selective attention studies compared to working memory studies (χ^2 (2) = 6.97, $P = 0.03$), as evidenced by a higher specificity ($\beta = 1.15$, $z = 2.96$, $P = 0.003$) but no significant difference in sensitivity ($\beta = 0.43$, $z = 1.08$, $P = 0.28$) (Fig. 3). In addition, the analysis revealed a significant overall effect of domain, with studies on the attention domain (K = 15) demonstrating higher specificity than those on the memory domain (K = 9, $\beta = 0.80$, $Z = 2.07$, $P = 0.04$) (Please see Table 2 for more details).

MRI Modality

When comparing the classification performances of t-fMRI studies across all neuropsychological domains with those of sMRI and r-fMRI studies, as reported in Kambeitz *et al.*'s meta-analysis,[3] no statistically significant differences were observed between t-fMRI and sMRI studies (ps >0.19) or between t-fMRI and r-fMRI studies (ps >0.56).

Notably, the bivariate meta-analytic model revealed that t-fMRI studies on the selective attention domain had better classification performances than sMRI studies (Bivariate model effect: χ^2 (2) = 14.89, $P < 0.001$; Sensitivity: $\beta = 0.87$, $z = 3.77$, $P < 0.001$; Specificity: $\beta = 0.76$, $z = 3.24$, $P = 0.001$) and r-fMRI studies (Bivariate model effect: χ^2 (2) = 8.80, $P = 0.01$; Specificity: $\beta = 0.79$, $z = 3.38$, $P = 0.001$; but not for Sensitivity: $\beta = 0.26$, $z = 1.07$, $P = 0.28$). However, we did not find any significant difference in sensitivity (P >0.30) or specificity (P >0.15) between neuroimaging modalities in studies on the working memory domain (see Fig. S2).

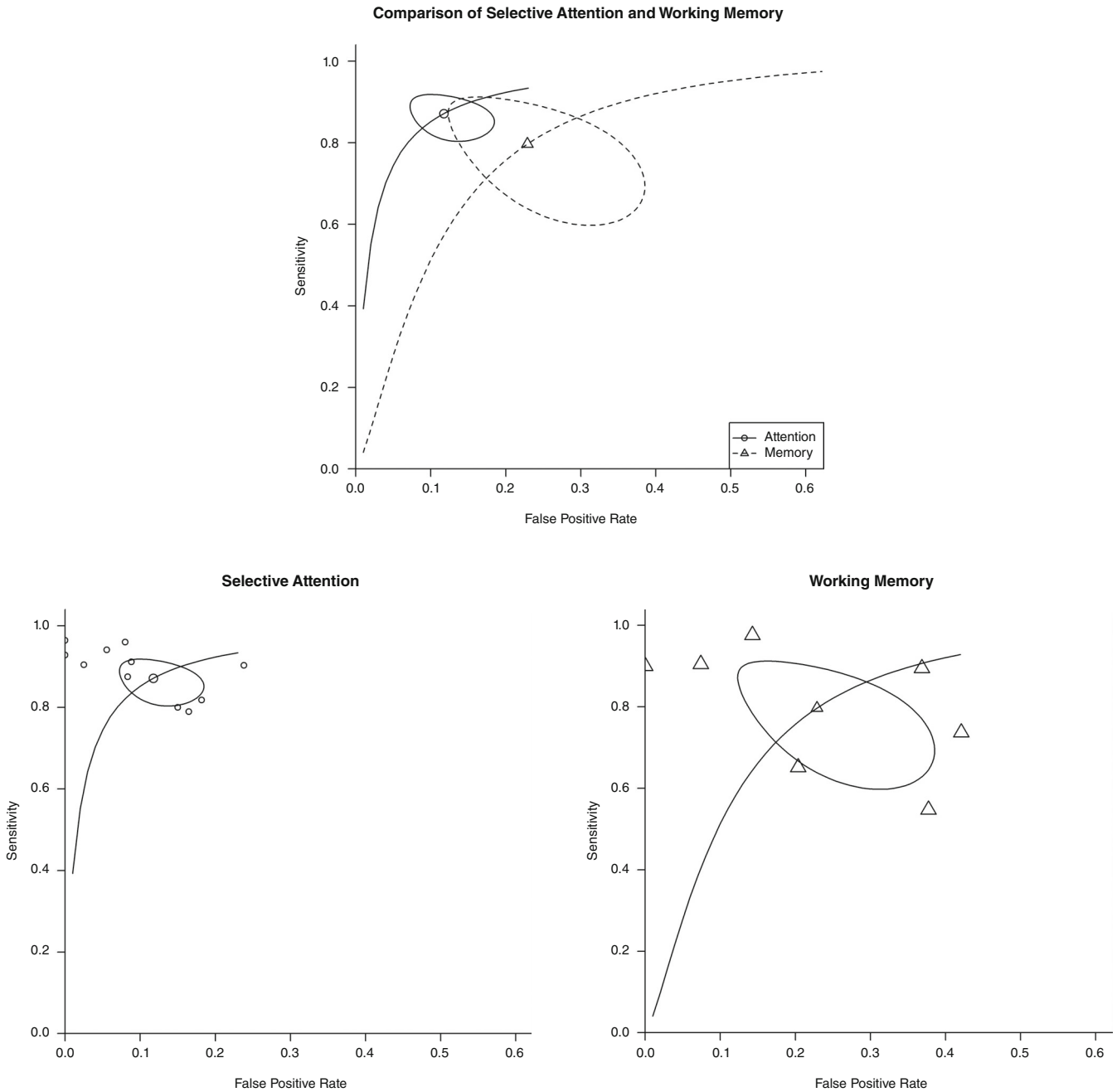

Fig. 3 Comparison of selective attention and working memory domains with summary receiver operating characteristic (sROC) curves. Selective attention and working memory domains are depicted separately (bottom).

Moderation effect of demographic, clinical, and ML approach-related factors

Demographic factors

Age

Our bivariate moderator analysis revealed a significant effect of age on classification performance (χ^2 (2) = 9.36, $P < 0.01$), indicating that older participants (including both HC and SCZ) generally exhibited higher sensitivity ($\beta = 0.10$, $z = 2.45$, $P = 0.01$) and specificity ($\beta = 0.08$, $z = 3.64$, $P < 0.001$) than younger participants (See Fig. 5).

Sex ratio

We did not find any significant effect of the proportions of males among all participants on sensitivity ($\beta = 2.37$, $z = 1.08$, $P = 0.28$) and specificity ($\beta = 0.60$, $z = 0.33$, $P = 0.74$). Further analyses showed that, neither the proportions of males among SCZ (sensitivity: $\beta = 2.25$, $z = 1.20$, $P = 0.23$; specificity: $\beta = 0.50$, $z = 0.35$, $P = 0.73$) nor that among HC (sensitivity: $\beta = 0.77$, $z = 0.40$, $P = 0.69$; specificity: $\beta = 1.75$, $z = 1.36$, $P = 0.17$) would have significant effect on sensitivity or specificity.

Clinical factors

Illness stage

When the studies enrolling chronic patients were separately analyzed, the meta-analysis showed a high sensitivity of 0.85 (95% CI: 0.77–0.91) and specificity of 0.87 (95% CI: 0.84–0.90). In contrast, studies recruiting first-episode psychosis (FEP) or high-risk (HR)

Table 2. Summarized classification performances of ML algorithm across neuropsychological domains

Neuropsychological domain	Research Number	Log Odds Ratio	Sensitivity		Specificity		Attention *vs.* Memory	Selective attention *vs.* Working memory
			Value	CI			Value	CI
Attention	15	3.28	0.81	0.72–0.88	0.86	0.81–0.91	χ^2 (2) = 4.01,	**χ^2 (2) = 6.97,**
Selective attention	11	3.89	0.87	0.82–0.91	0.88	0.83–0.92	$P = 0.135$	**$P = 0.03$**
Memory	9	2.47	0.79	0.67–0.88	0.76	0.61–0.87	Sensitivity:	Sensitivity:
Working memory	7	2.57	0.84	0.71–0.91	0.72	0.57–0.84	$\beta = 0.12$,	$\beta = 0.43$,
Emotion/Social cognition	4	3.53	0.88	0.83–0.91	0.83	0.76–0.88	$Z = 0.28$, $P = 0.78$	$Z = 1.08$, $P = 0.28$
Sensory-motor/ Perception	2	4.08	0.93	0.85–0.97	0.82	0.74–0.88	**Specificity:**	**Specificity:**
Language	1	5.54	0.89	0.75–0.96	0.92	0.79–0.97	**$\beta = 0.80$, $Z = 2.07$, $P = 0.04$**	**$\beta = 1.15$, $Z = 2.96$, $P = 0.003$**

Note: In table 2, we mentioned 3 p values for each subgroup analysis. Take "Attention *vs*. Memory" for example, the first *p* value after chi-squared indicates whether the fit of the model with Domain (2 levels: Attention or Memory) as the predictor is significantly better than the intercept-only model. The second and third *p* values indicate whether the corresponding beta coefficient is significantly different from zero. Those *p* values are the only 3 relevant *p* values resulted from the bivariate regression model.
Abbreviation: CI, Confidence interval.

individuals yielded a more modest sensitivity (0.63, 95% CI:0.50–0.74) and specificity (0.67, 95% CI:0.57–0.76). Furthermore, when the factor of "Illness stage" was added as a moderator variable to the bivariate meta-analysis model, the results revealed a significant difference in classification performance between the two illness stages (χ^2 (2) = 20.39, $P < 0.001$). The analysis revealed that chronic patients had a significantly higher sensitivity ($\beta = 1.15$, $z = 2.73$, $P = 0.006$) and specificity ($\beta = 1.21$, $z = 5.42$, $P < 0.001$) compared to FEP/HR individuals (see Fig. 4).

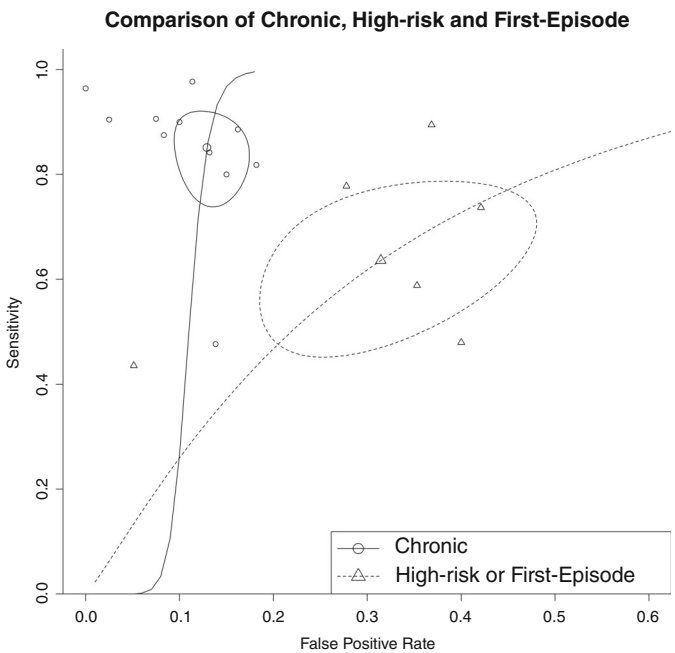

Fig. 4 Comparison of Chronic and High-risk/First-episode populations with summary receiver operating characteristic (sROC) curves.

Positive and negative symptoms

Relative scores of negative symptoms had an effect on classification performance, whereby more severe negative symptoms predicted higher specificity ($\beta = 6.03$, $z = 2.16$, $P = 0.031$) but not sensitivity ($\beta = 3.62$, $z = 0.99$, $P = 0.32$). In terms of converted scores, while we observed a similar trend in the moderating effect of negative symptoms on specificity, the results did not reach statistical significance ($P = 0.19$).

Conversely, relative scores of positive symptoms did not have any significant effect on sensitivity ($\beta = 2.67$, $z = 0.66$, $P = 0.51$) or specificity ($\beta = 1.41$, $z = 0.40$, $P = 0.69$) (See Fig. 5). Similarly, no moderating effect of converted scores of positive symptoms on sensitivity and specificity was observed ($P > 0.70$).

Medication dosage

There was no significant effect of medication dosage on either sensitivity or specificity ($P > 0.50$).

ML approach-related factors

Classification algorithm

Among the 31 studies included in our analysis, the SVM algorithm was most commonly used, and applied in 24 studies. Three studies used Neighbor-based algorithm (NBA), Random forest (RF), LGR, or LDA, while two studies applied the projection pursuit algorithm (PPA). The remaining algorithms, including ν-Multiple Kernel Analysis (v-MKL), Gaussian Naïve Bayes (GNB), Naïve Bayes (NB), Deep Learning (DL), Decision tree (DT), and K-Nearest Neighbor (KNN), were used in only one study each. However, due to the limited number of studies (<5) that utilized these other algorithms, we did not conduct moderator analysis, but instead presented the descriptive classification performance for each algorithm (See Table S2).

Features selected

The bivariate meta-analytic model showed no difference in sensitivity ($\beta = 0.67$, $z = 1.70$, $P = 0.09$) or specificity ($\beta = 0.18$, $z = 0.50$, $P = 0.62$) between studies that utilized functional connectivity/network properties (K = 8) and those that utilized voxel-based brain activation (K = 22) as features.

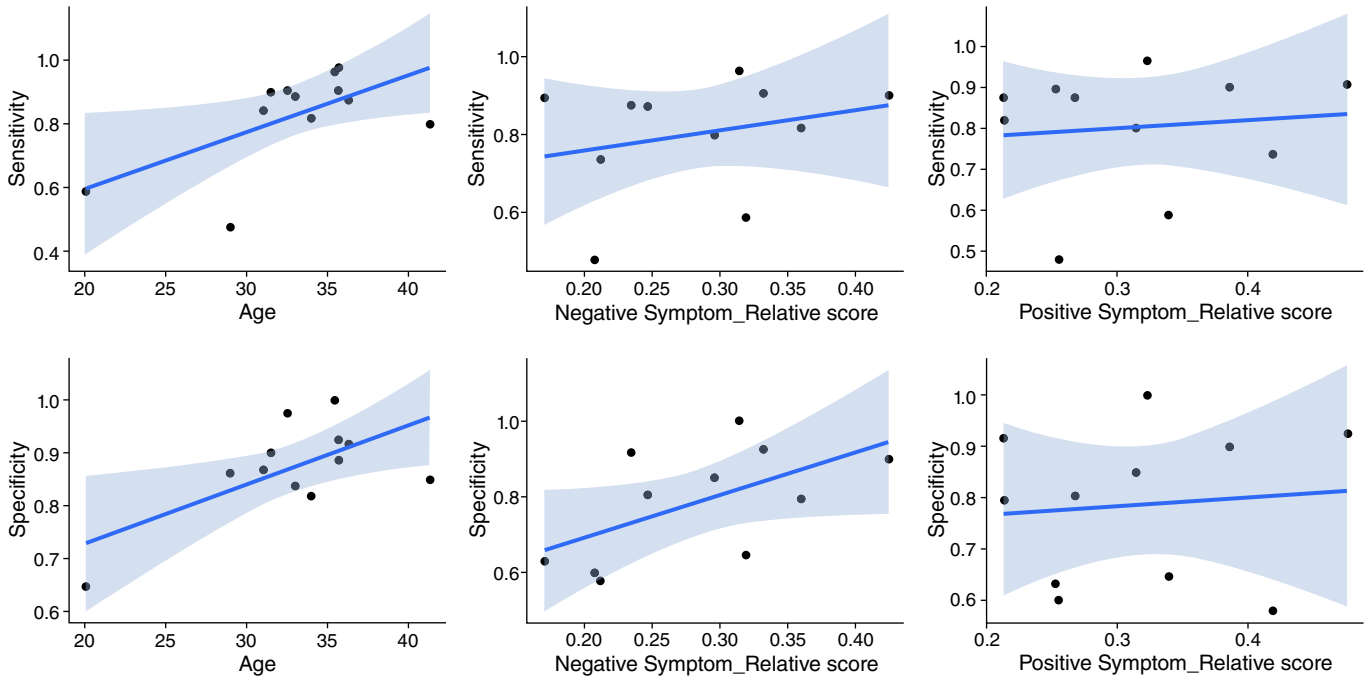

Fig. 5 Moderation effect of age, negative symptoms and positive symptoms on the diagnostic classification performances of schizophrenia.

Sample size and data source

Our bivariate meta-analytic model revealed that neither sample size nor data source had a significant effect on the classification performance (Sample size: χ^2 (2) = 2.10, $P = 0.35$; Data source: χ^2 (2) = 2.00, $P = 0.37$) (See Table S3).

Discussion

In this study, we utilized bivariate meta-analysis modeling to evaluate task-based neuroimaging phenotypes for classifying SCZ patients and HC. We analyzed 31 ML algorithm studies and found a sensitivity of 0.83 and specificity of 0.82 for neuropsychology-related brain functional responses. Studies on selective attention achieved higher specificity than studies on working memory. T-fMRI studies on selective attention had higher sensitivity and specificity than sMRI studies and higher specificity than r-fMRI studies. Illness stage and age had significant effects on sensitivity and specificity, while negative symptom severity had a significant effect on specificity.

Classification performances of the t-fMRI modality

Interestingly, our study found that t-fMRI studies on selective attention achieved higher specificity than studies on working memory, or rfMRI or sMRI studies. This suggests that classification performance may depend on the neuropsychological domains and neuroimaging modality. One plausible explanation is that selective attention is considered a core deficit in SCZ. It plays a pivotal role in modulating and optimizing other cognitive functioning, including perception, memory, and executive control. In essence, working memory determines which information is allowed to get access to conscious awareness, and thus to be encoded into memory, and to be processed subsequently. When such crucial function of selective attention is disrupted, a wide array of cognitive dysfunctions would develop, as in patients with SCZ.[43,44] Furthermore, impairments in selective attention may be related to dopaminergic dysregulation in the mesolimbic pathways.[45] SCZ is associated with dopamine dysregulation, and excessive dopaminergic transmission or the influence of dopamine antagonists in regions such as the ventral tegmental area of the midbrain can hinder the filtering of irrelevant information.[46] Failure in filtering irrelevant stimuli is thought to contribute to positive symptoms of SCZ, including delusion, hallucinations and disorganized thinking.[47,48] Therefore, investigating selective attention is important in SCZ, and may contribute to the underlying neuroimaging markers for precise diagnostic accuracy.

Besides, we found that ML classification algorithms had performed well in summarizing emotional/social cognition domains with a sensitivity of 0.88 and a specificity of 0.83. To foster precise evaluation of negative symptoms of SCZ, it is imperative to assess the efficacy of ML classification algorithms, encompassing not only attention and memory domains but also domains like emotion, reward processing, and social cognition in alignment with the Positive Valence System of the RDoC framework.[49]

Effect of clinical factors

The bivariate meta-analytic model revealed that studies on chronic SCZ patients had higher sensitivity and specificity than those on first-episode patients and high-risk individuals. Several reasons may explain this finding. First, long DOI is associated with more severe alterations in brain structure and function in SCZ,[50,51] which is consistent with both the neurodevelopmental and neurodegenerative theories of SCZ[52,53]. For instance, a 10-year follow-up study found that patients with SCZ exhibited a progressive decline in intelligence and specific neurocognitive domains, including verbal knowledge and memory, following the onset of the illness.[54] Furthermore, reductions in cortical thickness in frontal and temporal brain areas have been observed prior to the onset of psychosis.[55] The SCZ-associated structural and functional abnormalities progressed after onset of the illness[56,57] and are associated with clinical outcomes and cognitive impairments.[58–61] While the prominent neural alterations of chronic SCZ enhance the classification model, such models may limit the generalizability of our findings to patients with first-onset SCZ or the ultra-high risk populations. More effort is needed to find an optimal classification algorithm and sensitive domain to improve the diagnosis or detection of first-episode patients and individuals at high risk of psychosis. Future studies should use larger samples of young patients with chronic SCZ to disentangle the effects of chronicity and age of patients with SCZ. The issue that chronicity could be confounded with age, and the moderation effect of age should be addressed.[3]

Additionally, our findings suggest that the severity of negative symptoms, rather than positive symptoms, might be a potential predictor of higher specificity in our classification results. This is consistent with existing evidence that demonstrates a link between dysregulated cortico-striatal networks and hypodopaminergic functioning in prefrontal and mesolimbic regions with negative symptoms in SCZ patients.[62–64] Negative symptoms can manifest early in the course of illness[65] and persist for a longer time than positive symptoms.[4,63,66] In fact, the severity of negative symptoms in patients with first-episode psychosis predicted poor functional outcomes 1 year later.[67,68] These findings suggest that the neural correlates of persistent negative symptoms may be more distinct than those of positive symptoms, leading to higher classification performance.

We did not find significant effect of medication. However, higher doses of antipsychotic medications, particularly first-generation antipsychotics (FGA), are associated with more significant gray matter loss in the prefrontal and temporal lobes, regions linked to cognitive impairment and functional deficits.[69] A recent meta-analysis examining structural and resting-state functional MRI data also identified a significant impact of antipsychotic medication on specificity, with higher specificity observed in SCZ patients receiving higher antipsychotic doses.[3] Thus, antipsychotic exposure likely influences machine learning classification performance based on t-fMRI data.[70] Although the present meta-analysis did not detect a moderating effect of medication dosage, FGAs and second-generation antipsychotics (SGAs) may exhibit distinct profiles of dopamine D2 receptor affinity, which can differentially affect mesolimbic and prefrontal brain regions, as well as the related cognitive and emotional functions.[71] Therefore, future research should systematically investigate whether and how different types of antipsychotic medications can influence machine learning classification performance in the context of t-fMRI data.

The role of the ML classification algorithm

Similar to meta-analytic findings from sMRI and rfMRI studies,[3] the SVM algorithm is the most frequently used ML approach in t-fMRI studies, applied in 24 (77%) studies. The SVM algorithm is particularly effective in dealing with imaging data in high-dimensional spaces, making it a suitable choice for small datasets in t-fMRI studies. However, no conclusive remark could be made favoring any ML approach due to the limited number of studies utilizing algorithms other than SVM. For instance, three studies[6,72,73] that applied LDA to t-fMRI data reported varying sensitivities (0.59–0.90) and specificities (0.64–0.88), while three other independent studies using LGR[22,72] demonstrated substantial classification performance heterogeneity concerning the ML algorithm (Sensitivity: 0.60–0.90; Specificity: 0.47–0.74). Only one study applied the DL algorithm, which may reduce the artificial influence by extracting less redundant and more informative features, potentially overcoming the limitations of the SVM algorithm. The DL algorithm has been applied to the diagnostic modeling of SCZ in the sMRI and rfMRI modalities.[74–76] Nevertheless, its effectiveness in the t-fMRI modality and the reasons underlying the heterogeneity of ML algorithms remain unclear. Therefore, a systematic comparison of different classification algorithms for t-fMRI-based classification of SCZ is necessary to address these concerns.

Despite the prevailing notion that SCZ is a disorder of brain network organization rather than individual voxel-based activation in specific brain areas,[77–79] recent studies have shown that using functional connectivity as features does not necessarily lead to better classification performance compared to voxel-based regional activation. However, it is noteworthy that comparing different selected features in the context of meta-analysis could be confounded by neuropsychological domains. In particular, nine out of 22 studies that used voxel-based activation as features measured the domain of selective attention, resulting in a higher sensitivity and specificity than studies that used functional connection as features (one out of eight studies). Interestingly, three studies that incorporated brain network properties (e.g. degree, clustering coefficient, etc.) as features in the diagnostic model yielded a much higher sensitivity of 0.91 and specificity of 0.88. Despite these findings, a systematic investigation of machine learning algorithms with brain network properties as features for domain-specific t-fMRI-based classification of SCZ is still lacking and needs further exploration.

In this study, we did not observe any significant effect of sample size on classification performance. Nevertheless, sample size remains an essential factor in machine learning-based classification models. Majority of classification models included in this study were built on relatively small samples, compared to structural or resting-state fMRI. Kalmady *et al.*[80] have further demonstrated that smaller sample sizes in single-site studies yielded higher prediction accuracy for SCZ using neuroimaging data. Future studies would benefit from larger sample sizes, multiple-site validation as well as the implementation of optimized algorithms (e.g. ensemble models that combine predictions from multiple 'single-source' models, as suggested by Kalmady). These steps are important to strengthen the reliability and validity of the results.[81]

Limitations

Some limitations should be acknowledged. First, most of the studies had small sample sizes and did not include a cross-validation sample. Second, the majority of studies focused solely on distinguishing patients with SCZ from healthy controls. It remains uncertain how effective the classification model would be in distinguishing SCZ from other mental illnesses such as bipolar disorder, and obsessive-compulsive disorder.[19,23,82] Third, it is crucial to acknowledge the problem of publication bias in studies on machine learning algorithms.[81] Our analysis of the funnel plot of published studies revealed a significant asymmetry, indicating that smaller and negative studies may be underrepresented. Several factors could contribute to publication bias, including biased classification performance in earlier years. As our analysis demonstrated, there was a significant negative trend between specificity and the year of publication ($\beta = -0.069$, $z = -1.91$, $P = 0.056$), possibly attributable to variations in the utility standard of machine learning algorithms over time. Finally, our methodology of utilizing the best classification performance of each study, in cases where multiple ML algorithms were employed, may have led to overestimating the effectiveness of ML algorithms in diagnosing SCZ. Therefore, it is crucial to interpret the findings of our study in light of potential publication biases, and to exercise caution in generalizing the results.

Conclusions

Our meta-analysis provided evidence that ML techniques can effectively classify patients with SCZ using t-fMRI data, albeit potential publication bias. Furthermore, our study identified several key factors, including illness stage, age, severity of negative symptoms, and neuropsychological domain that could influence the effectiveness of ML-based classification models for SCZ. However, it is important to acknowledge that the application of ML algorithms in this context is still in its early stage. The performance of these algorithms can be influenced by various factors, including the choice of algorithm, parameter tuning, data quality and quantity, as well as issues related to validation and generalization. However, more work is needed before we can integrate ML into clinical practice.

Author contributions

Xuan Wang: Formal analysis, Visualization, Writing – original draft. Chao Yan: Conceptualization, Writing – original draft, Writing – Review & Editing, Funding acquisition. Peng-yuan Yang: Formal analysis, Visualization, Validation. Zheng Xia: Validation. Xin-lu Cai: Resources. Yi Wang: Writing – Review & Editing. Sze Chai Kwok: Writing – Review & Editing. Raymond C.K. Chan: Writing – Review & Editing.

Acknowledgements

This study was supported by grants from the National Natural Science Foundation of China [32171084], the Natural Science Foundation of Shanghai [21ZR1421000] and the MOE (Ministry of Education of China) Project of Humanities and Social Sciences [20YJC190025]. Raymond Chan was supported by the STI2030-Major Projects (2021ZD0200800) and the Philip K. H. Foundation. These funding agents had no role in the study design, data analysis, and interpretation; or writing of the manuscript.

Disclosure statement

The authors have no conflicts of interest to declare.

References

1. Tandon R, Gaebel W, Barch DM *et al*. Definition and description of schizophrenia in the DSM-5. *Schizophr. Res.* 2013; **150**: 3–10.
2. Janssen RJ, Mourão-Miranda J, Schnack HG. Making individual prognoses in psychiatry using neuroimaging and machine learning. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 2018; **3**: 798–808.
3. Kambeitz J, Kambeitz-Ilankovic L, Leucht S *et al*. Detecting neuroimaging biomarkers for schizophrenia: A meta-analysis of multivariate pattern recognition studies. *Neuropsychopharmacology* 2015; **40**: 1742–1751.
4. Schnack HG. Improving individual predictions: Machine learning approaches for detecting and attacking heterogeneity in schizophrenia (and other psychiatric diseases). *Schizophr. Res.* 2019; **214**: 34–42.
5. Orban P, Dansereau C, Desbois L *et al*. Multisite generalizability of schizophrenia diagnosis classification based on functional brain connectivity. *Schizophr. Res.* 2018; **192**: 167–171.
6. 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**: 252–261.
7. Modinos G, Pettersson-Yeo W, Allen P, McGuire PK, Aleman A, Mechelli A. Multivariate pattern classification reveals differential brain activation during emotional processing in individuals with psychosis proneness. *Neuroimage* 2012; **59**: 3033–3041.
8. Valli I, Marquand AF, Mechelli A *et al*. Identifying individuals at high risk of psychosis: Predictive utility of support vector machine using structural and functional MRI data. *Front. Psych.* 2016; **7**: 52.
9. Iwabuchi SJ, Liddle PF, Palaniyappan L. Clinical utility of machine-learning approaches in schizophrenia: Improving diagnostic confidence for translational neuroimaging. *Front. Psych.* 2013; **4**: 95.
10. Sun D, van Erp TG, Thompson PM *et al*. Elucidating a magnetic resonance imaging-based neuroanatomic biomarker for psychosis: Classification analysis using probabilistic brain atlas and machine learning algorithms. *Biol. Psychiatry* 2009; **66**: 1055–1060.
11. Ardekani BA, Tabesh A, Sevy S, Robinson DG, Bilder RM, Szeszko PR. Diffusion tensor imaging reliably differentiates patients with schizophrenia from healthy volunteers. *Hum. Brain Mapp.* 2011; **32**: 1–9.
12. Caan MW, Vermeer KA, van Vliet LJ *et al*. Shaving diffusion tensor images in discriminant analysis: A study into schizophrenia. *Med. Image Anal.* 2006; **10**: 841–849.
13. Ota M, Sato N, Ishikawa M *et al*. Discrimination of female schizophrenia patients from healthy women using multiple structural brain measures obtained with voxel-based morphometry. *Psychiatry Clin. Neurosci.* 2012; **66**: 611–617.
14. Kaufmann T, Skåtun KC, Alnæs D *et al*. Disintegration of sensorimotor brain networks in schizophrenia. *Schizophr. Bull.* 2015; **41**: 1326–1335.
15. Yu Y, Shen H, Zhang H, Zeng LL, Xue Z, Hu D. Functional connectivity-based signatures of schizophrenia revealed by multiclass pattern analysis of resting-state fMRI from schizophrenic patients and their healthy siblings. *Biomed. Eng. Online* 2013; **12**: 10.
16. 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**: 1265–1275.
17. Kiehl KA, Liddle PF. An event-related functional magnetic resonance imaging study of an auditory oddball task in schizophrenia. *Schizophr. Res.* 2001; **48**: 159–171.
18. Walter H, Kammerer H, Frasch K, Spitzer M, Abler B. Altered reward functions in patients on atypical antipsychotic medication in line with the revised dopamine hypothesis of schizophrenia. *Psychopharmacology (Berl)* 2009; **206**: 121–132.
19. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. *Annu. Rev. Clin. Psychol.* 2018; **14**: 91–118.
20. Demirci O, Clark VP, Magnotta VA *et al*. A review of challenges in the use of fMRI for disease classification/characterization and a projection pursuit application from multi-site fMRI schizophrenia study. *Brain Imaging Behav.* 2008; **2**: 147–226.
21. Bendfeldt K, Smieskova R, Koutsouleris N *et al*. Classifying individuals at high-risk for psychosis based on functional brain activity during working memory processing. *Neuroimage. Clin.* 2015; **9**: 555–563.
22. Ghosal S, Chen Q, Pergola G *et al*. A generative-discriminative framework that integrates imaging, genetic, and diagnosis into coupled low dimensional space. *Neuroimage* 2021; **238**: 118200.
23. Arbabshirani MR, Plis S, Sui J, Calhoun VD. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage* 2017; **145**: 137–165.
24. Góngora Alonso S, Marques G, Agarwal D, De la Torre DI, Franco-Martín M. Comparison of machine learning algorithms in the prediction of hospitalized patients with schizophrenia. *Sensors (Basel)* 2022; **22**: 2517.
25. Yassin W, Nakatani H, Zhu Y *et al*. Machine-learning classification using neuroimaging data in schizophrenia, autism, ultra-high risk and first-episode psychosis. *Transl. Psychiatry* 2020; **10**: 278.
26. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J. Clin. Epidemiol.* 2005; **58**: 982–990.
27. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med.* 2009; **6**: e1000097.
28. Van Houwelingen HC, Zwinderman KH, Stijnen T. A bivariate approach to meta-analysis. *Stat. Med.* 1993; **12**: 2273–2284.
29. Juneja A, Rana B, Agrawal RK. A novel approach for classification of schizophrenia patients and healthy subjects using auditory oddball functional MRI. **In:** *2014 13th Mexican International Conference on Artificial Intelligence*, Tuxtla Gutierrez, Mexico, 2014; 75–81Juneja A.
30. Juneja A, Rana B, Agrawal RK. A combination of singular value decomposition and multivariate feature selection method for diagnosis of schizophrenia using fMRI. *Biomed. Signal Process. Control* 2016; **27**: 122–133.
31. Juneja A, Rana B, Agrawal RK. A novel fuzzy rough selection of non-linearly extracted features for schizophrenia diagnosis using fMRI. *Comput. Methods Programs Biomed.* 2018; **155**: 139–152.
32. Juneja A, Rana B, Agrawal RK. fMRI based computer aided diagnosis of schizophrenia using fuzzy kernel feature extraction and hybrid feature selection. *Multimed. Tools Appl.* 2018; **77**: 3963–3989.
33. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J. Clin. Epidemiol.* 2005; **58**: 882–893.
34. Orwin RG. A fail-SafeN for effect size in meta-analysis. *J. Educ. Stat.* 1983; **8**: 157–159.
35. Rosenthal R. The file drawer problem and tolerance for null results. *Psychol. Bull.* 1979; **86**: 638–641.
36. van Erp TGM, Preda A, Nguyen D *et al*. Converting positive and negative symptom scores between PANSS and SAPS/SANS. *Schizophr. Res.* 2014; **152**: 289–294.
37. Fava GA, Kellner R. Staging: A neglected dimension in psychiatric classification. *Acta Psychiatr. Scand.* 1993; **87**: 225–230.
38. Fusar-Poli P, Borgwardt S, Bechdolf A *et al*. The psychosis high-risk state: A comprehensive state-of-the-art review. *JAMA Psychiatry* 2013; **70**: 107–120.
39. Anda L, Brønnick KK, Johannessen JO *et al*. Cognitive profile in ultra high risk for psychosis and schizophrenia: A comparison using coordinated norms. *Front. Psych.* 2019; **10**: 695.
40. Benetti S, Mechelli A, Picchioni M, Broome M, Williams S, McGuire P. Functional integration between the posterior hippocampus and prefrontal cortex is impaired in both first episode schizophrenia and the at risk mental state. *Brain* 2009; **132**: 2426–2436.
41. Doebler P, Holling H. Meta-Analysis of Diagnostic Accuracy with Mada. 2017.
42. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res. Synth. Methods* 2010; **1**: 97–111.
43. Braff DL. Information processing and attention dysfunctions in schizophrenia. *Schizophr. Bull.* 1993; **19**: 233–259.
44. Luck SJ, Gold JM. The construct of attention in schizophrenia. *Biol. Psychiatry* 2008; **64**: 34–39.
45. Nieoullon A. Dopamine and the regulation of cognition and attention. *Prog. Neurobiol.* 2002; **67**: 53–83.

46. Millard SJ, Bearden CE, Karlsgodt KH, Sharpe MJ. The prediction-error hypothesis of schizophrenia: New data point to circuit-specific changes in dopamine activity. *Neuropsychopharmacology* 2022; **47**: 628–640.
47. Lubow RE, Weiner I, Schlossberg A, Baruch I. Latent inhibition and schizophrenia. *Bull. Psych. Soc.* 1987; **25**: 464–467.
48. Weiner I, Arad M. Using the pharmacology of latent inhibition to model domains of pathology in schizophrenia and their treatment. *Behav. Brain Res.* 2009; **204**: 369–386.
49. Insel T, Cuthbert B, Garvey M *et al*. Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *Am. J. Psychiatry* 2010; **167**: 748–751.
50. Hashimoto N, Ito YM, Okada N *et al*. The effect of duration of illness and antipsychotics on subcortical volumes in schizophrenia: Analysis of 778 subjects. *NeuroImage-Clin.* 2018; **17**: 563–569.
51. Shen C-L, Tsai S-J, Lin C-P, Yang ACC. Progressive brain abnormalities in schizophrenia across different illness periods: A structural and functional MRI study. *Schizophrenia (Heidelb)* 2023; **9**: 2.
52. Rapoport JL, Giedd JN, Gogtay N. Neurodevelopmental model of schizophrenia: Update 2012. *Mol. Psychiatry* 2012; **17**: 1228–1238.
53. Pérez-Neri I, Ramírez-Bermúdez J, Montes S, Ríos C. Possible mechanisms of neurodegeneration in schizophrenia. *Neurochem. Res.* 2006; **31**: 1279–1294.
54. Zanelli J, Mollon J, Sandin S *et al*. Cognitive change in schizophrenia and other psychoses in the decade following the first episode. *Am. J. Psychiatry* 2019; **176**: 811–819.
55. Mechelli A, Riecher-Rössler A, Meisenzahl EM *et al*. Neuroanatomical abnormalities that predate the onset of psychosis: A multicenter study. *Arch. Gen. Psychiatry* 2011; **68**: 489–495.
56. Cao H, Chung Y, McEwen SC *et al*. Progressive reconfiguration of resting-state brain networks as psychosis develops: Preliminary results from the north American Prodrome longitudinal study (NAPLS) consortium. *Schizophr. Res.* 2020; **226**: 30–37.
57. van Haren NE, Cahn W, Hulshoff Pol HE, Kahn RS. The course of brain abnormalities in schizophrenia: Can we slow the progression? *J. Psychopharmacol.* 2012; **26**: 8–14.
58. Antonucci LA, Pergola G, Pigoni A *et al*. A pattern of cognitive deficits stratified for genetic and environmental risk reliably classifies patients with schizophrenia from healthy control subjects. *Biol. Psychiatry* 2020; **87**: 697–707.
59. Nielsen M, Rostrup E, Broberg BV, Wulff S, Glenthøj B. Negative symptoms and reward disturbances in schizophrenia before and after antipsychotic monotherapy. *Clin. EEG Neurosci.* 2018; **49**: 36–45.
60. Vives-Gilabert Y, Abdulkadir A, Kaller CP *et al*. Detection of preclinical neural dysfunction from functional connectivity graphs derived from task fMRI. An example from degeneration. *Psychiatry Res.* 2013; **214**: 322–330.
61. Wu R, Ou Y, Liu F *et al*. Reduced brain activity in the right putamen as an early predictor for treatment response in drug-naive, first-episode schizophrenia. *Front. Psychiatry* 2019; **10**: 741.
62. Abi-Dargham A. Do we still believe in the dopamine hypothesis? New data bring new evidence. *Int. J. Neuropsychopharmacol.* 2004; **7**: S1–S5.
63. Correll CU, Schooler NR. Negative symptoms in schizophrenia: A review and clinical guide for recognition, assessment, and treatment. *Neuropsychiatr. Dis. Treat.* 2020; **16**: 519–534.
64. Millan MJ, Fone K, Steckler T, Horan WP. Negative symptoms of schizophrenia: Clinical characteristics, pathophysiological substrates, experimental models and prospects for improved treatment. *Eur. Neuropsychopharmacol.* 2014; **24**: 645–692.
65. Valmaggia LR, Stahl D, Yung AR *et al*. Negative psychotic symptoms and impaired role functioning predict transition outcomes in the at-risk mental state: A latent class cluster analysis study. *Psychol. Med.* 2013; **43**: 2311–2325.
66. Chang WC, Hui CL, Tang JY *et al*. Persistent negative symptoms in first-episode schizophrenia: A prospective three-year follow-up study. *Schizophr. Res.* 2011; **133**: 22–28.
67. Bergé D, Mané A, Salgado P *et al*. Predictors of relapse and functioning in first-episode psychosis: A two-year follow-up study. *Psychiatr. Serv.* 2016; **67**: 227–233.
68. Hovington CL, Bodnar M, Joober R, Malla AK, Lepage M. Identifying persistent negative symptoms in first episode psychosis. *BMC Psychiatry* 2012; **12**: 224.
69. Kubota M, van Haren NEM, Haijma SV *et al*. Association of IQ changes and progressive brain changes in patients with schizophrenia. *JAMA Psychiatry* 2015; **72**: 803–812.
70. Zeng J, Zhang W, Wu G *et al*. Effects of antipsychotic medications and illness duration on brain features that distinguish schizophrenia patients. *Schizophr. Bull.* 2022; **48**: 1354–1362.
71. Yang C, Tang J, Liu N *et al*. The effects of antipsychotic treatment on the brain of patients with first-episode schizophrenia: A selective review of longitudinal MRI studies. *Front. Psych.* 2021; **12**: 593703.
72. Potvin S, Giguère C, Mendrek A. Functional connectivity during visuospatial processing in schizophrenia: A classification study using Lasso regression. *Neuropsychiatr. Dis. Treat.* 2021; **17**: 1077–1087.
73. Yoon JH, Nguyen DV, McVay LM *et al*. Automated classification of fMRI during cognitive control identifies more severely disorganized subjects with schizophrenia. *Schizophr. Res.* 2012; **135**: 28–33.
74. Hashimoto Y, Ogata Y, Honda M, Yamashita Y. Deep feature extraction for resting-state functional MRI by self-supervised learning and application to schizophrenia diagnosis. *Front. Neurosci.* 2021; **15**: 696853.
75. 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.
76. Zheng J, Wei X, Wang J, Lin H, Pan H, Shi Y. Diagnosis of schizophrenia based on deep learning using fMRI. *Comput. Math. Methods Med.* 2021; **2021**: 8437260.
77. Friston K, Brown HR, Siemerkus J, Stephan KE. The dysconnection hypothesis (2016). *Schizophr. Res.* 2016; **176**: 83–94.
78. Rubinov M, Bullmore E. Schizophrenia and abnormal brain network hubs. *Dialogues Clin. Neurosci.* 2013; **15**: 339–349.
79. Whitfield-Gabrieli S, Ford JM. Default mode network activity and connectivity in psychopathology. *Annu. Rev. Clin. Psychol.* 2012; **8**: 49–76.
80. Kalmady SV, Greiner R, Agrawal R *et al*. Towards artificial intelligence in mental health by improving schizophrenia prediction with multiple brain parcellation ensemble-learning. *NPJ Schizophr.* 2019; **5**: 2.
81. Vieira S, Gong QY, Pinaya WHL *et al*. Using machine learning and structural neuroimaging to detect first episode psychosis: Reconsidering the evidence. *Schizophr. Bull.* 2020; **46**: 17–26.
82. Bzdok D, Meyer-Lindenberg A. Machine learning for precision psychiatry: Opportunities and challenges. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 2018; **3**: 223–230.
83. Demirci O, Clark VP, Calhoun VD. A projection pursuit algorithm to classify individuals using fMRI data: Application to schizophrenia. *Neuroimage* 2008; **39**: 1774–1782.
84. Sui J, Adali T, Pearlson GD, Calhoun VD. An ICA-based method for the identification of optimal FMRI features and components using combined group-discriminative techniques. *Neuroimage* 2009; **46**: 73–86.
85. 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.
86. Costafreda SG, Fu CH, Picchioni M *et al*. Pattern of neural responses to verbal fluency shows diagnostic specificity for schizophrenia and bipolar disorder. *BMC Psychiatry* 2011; **11**: 18.
87. Du W, Calhoun VD, Li H *et al*. High classification accuracy for schizophrenia with rest and task FMRI data. *Front. Hum. Neurosci.* 2012; **6**: 145.
88. Rish I, Cecchi G, Thyreau B *et al*. Schizophrenia as a network disease: Disruption of emergent brain function in patients with auditory hallucinations. *PloS One* 2013; **8**: e50625.
89. Castro E, Gómez-Verdejo V, Martínez-Ramón M, Kiehl KA, Calhoun VD. A multiple kernel learning approach to perform classification of groups from complex-valued fMRI data analysis: Application to schizophrenia. *Neuroimage* 2014; **87**: 1–17.
90. Mascolo L, Guccione P, Nico G, Taurisano P, Fazio L. Data driven analysis of functional brain networks in fMRI for schizophrenia investigation. *Int. J. Imaging Syst. Technol.* 2014; **24**: 239–248.
91. Gaebler AJ, Mathiak K, Koten JW *et al*. Auditory mismatch impairments are characterized by core neural dysfunctions in schizophrenia. *Brain* 2015; **138**: 1410–1423.
92. Koch SP, Hägele C, Haynes J-D, Heinz A, Schlagenhauf F, Sterzer P. Diagnostic classification of schizophrenia patients on the basis of regional reward-related FMRI signal patterns. *PloS One* 2015; **10**: e0119089.
93. Viviano JD, Buchanan RW, Calarco N *et al*. Resting-state connectivity biomarkers of cognitive performance and social function in individuals with schizophrenia Spectrum disorder and healthy control subjects. *Biol. Psychiatry* 2018; **84**: 665–674.
94. Lombardi A, Guaragnella C, Amoroso N *et al*. Modelling cognitive loads in schizophrenia by means of new functional dynamic indexes. *Neuroimage* 2019; **195**: 150–164.

95. Oh K, Kim W, Shen G *et al*. Classification of schizophrenia and normal controls using 3D convolutional neural network and outcome visualization. *Schizophr. Res.* 2019; **212**: 186–195.
96. Antonucci LA, Penzel N, Pergola G *et al*. Multivariate classification of schizophrenia and its familial risk based on load-dependent attentional control brain functional connectivity. *Neuropsychopharmacology* 2020; **45**: 613–621.
97. Yang H, Di X, Gong Q, Sweeney J, Biswal B. Investigating inhibition deficit in schizophrenia using task-modulated brain networks. *Brain Struct. Funct.* 2020; **225**: 1601–1613.
98. Cruz-Martinez C, Reyes-Garcia CA, Vanello N. A novel event-related fMRI supervoxels-based representation and its application to schizophrenia diagnosis. *Comput. Methods Programs Biomed.* 2022; **213**: 106509.

Supporting Information

Additional supporting information can be found online in the Supporting Information section at the end of this article.