

# Applying machine-learning and deep-learning to predict depression from brain MRI and identify depression-related brain biology

Received: 5 March 2025

Revised: 22 December 2025

Accepted: 30 January 2026

Cite this article as: Jiang, J.-C., Brianceau, C., Delzant, E. *et al.* Applying machine-learning and deep-learning to predict depression from brain MRI and identify depression-related brain biology. *Transl Psychiatry* (2026). <https://doi.org/10.1038/s41398-026-03889-8>

Jiayue-Clara Jiang, Camille Brianceau, Elise Delzant, Romain Colle, Hugo Bottemanne, Emmanuelle Corruble, Naomi R. Wray, Olivier Colliot, Sonia Shah & Baptiste Couvy-Duchesne

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

## **Applying Machine-Learning and Deep-Learning to Predict Depression from Brain MRI and Identify Depression-Related Brain Biology**

Jiayue-Clara Jiang (PhD)<sup>1\*</sup>, Camille Brianceau (MSc)<sup>2</sup>, Elise Delzant (PhD)<sup>2</sup>, Romain Colle (MD,PhD)<sup>3,4</sup>, Hugo Bottemanne (MD,PhD)<sup>3,4</sup>, Emmanuelle Corruble (MD,PhD)<sup>3,4</sup>, Naomi R Wray (PhD)<sup>1,5</sup>, Olivier Colliot (PhD)<sup>2</sup>, Sonia Shah (PhD)<sup>1</sup>, Baptiste Couvy-Duchesne (PhD)<sup>1,2,6,7</sup>

<sup>1</sup>Institute for Molecular Bioscience, The University of Queensland, St Lucia, QLD, Australia

<sup>2</sup>Sorbonne University, Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France

<sup>3</sup>Service Hospitalo-Universitaire de Psychiatrie de Bicêtre, Mood Center Paris Saclay, Assistance Publique-Hôpitaux de Paris, Hôpitaux Universitaires Paris-Saclay, Hôpital de Bicêtre, Le Kremlin Bicêtre, F-94275, France

<sup>4</sup>MOODS Team, INSERM 1018, CESP (Centre de Recherche en Epidémiologie et Santé des Populations), Université Paris-Saclay, Faculté de Médecine Paris-Saclay, Le Kremlin Bicêtre, F-94275, France

<sup>5</sup>Department of Psychiatry, University of Oxford, Oxford

<sup>6</sup>Brain & Mental Health Program, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia

<sup>7</sup>School of Biomedical Sciences, Queensland University of Technology, Brisbane, QLD, Australia

\*corresponding author

**Running title:** Brain-MRI-based prediction of depression

**Corresponding author:**

Jiayue-Clara Jiang, PhD

Address: 306 Carmody Rd, Institute for Molecular Bioscience, The University of Queensland, St Lucia, Australia, 4072

Phone: +61 7 3443 6062

Email: j.jiang@uq.edu.au

**Abstract**

The accuracy of grey-matter predictors of depression has remained limited. In this study, brain-based predictors of major depressive disorder (MDD) were trained using machine-learning (Best Linear Unbiased Predictors [BLUP]) and deep-learning (ResNet3D) techniques applied to high-dimensional (voxel-wise) grey-matter structure extracted from T1-weighted structural MRI. The training sample comprised 987 MDD cases and 3,934 controls from the UK Biobank. Predictors were evaluated in an independent sub-cohort of 483 MDD cases and 1,939 controls from the UK Biobank and replicated in a clinical cohort (DEP-ARREST CLIN) of 64 cases and 32 controls. In the UK Biobank, the BLUP predictor showed a significant association with MDD status (AUC=0.57; OR=1.28 [1.15-1.43]; p-value= $1.1 \times 10^{-5}$ ), which was confirmed in both males and females. By partitioning the BLUP predictor by brain regions of interest (ROI), we found nominal significance supporting the contribution of previously identified MDD-related ROIs (e.g. hippocampus and amygdala), though none passed multiple testing correction. The BLUP predictor overlapped partially with a polygenic score (PGS) of major depression (AUC=0.65) but also captured a nominally significant signal that was not captured by the genetic score (combined AUC=0.66, p-value=0.024 when compared to PGS alone). No association passed multiple testing correction in the DEP-ARREST CLIN cohort, likely due to the small sample size. In contrast, the deep-learning predictor was not associated with MDD after multiple testing corrections. We estimated the morphometricity of MDD to be 0.061, implying limited potential of a brain-based predictor based on grey-matter structure (maximal AUC=0.64). While the modest AUC values reiterate the challenge of developing brain-based MDD predictors for clinical applications, our predictors inform future research to explore brain-based relationships between MDD and comorbidities.

## Introduction

Despite being a leading contributor to the global burden of disease for people aged 10 to 74 years [1], depression remains poorly understood, possibly due to its heterogeneous presentation and multi-causal aetiology influenced by genetic, environmental and social factors [2]. Depression genetic risk loci identified from genome-wide association studies (GWAS) have been shown to be enriched for genes expressed in specific brain regions, specifically the anterior cingulate cortex and frontal cortex [3]. Furthermore, several brain structural features, such as larger hippocampal volumes and higher frontal cortical thickness, have been found to predict remission after antidepressant treatment [4]. These pieces of evidence converge to suggest a complex and dynamic implication of brain regions in leading to depression and over the course of depression episodes and treatment.

Brain imaging, such as Magnetic Resonance Imaging (MRI), has advanced the understanding of the brain structures and functions associated with depression. Historically, neuroimaging studies have been limited by small sample sizes, leading to few robust results [5, 6]. The Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) consortium has gathered brain images from thousands of depression cases and controls from more than 20 sites to perform large-scale analyses [7]. They consistently reported a reduced hippocampal volume amongst individuals with depression, whereas the findings on structural changes in other brain regions, such as the amygdala and thalamus, have remained conflicting [2]. The variability in results may be due to different imaging protocols and machines as well as the heterogeneous nature of depression, manifesting as different symptoms, severity and subtype categories. At the same time, the large-scale ENIGMA-led efforts to build brain-based predictors of depression have not proven successful, highlighting the difficulty of this task [8]. More recently, the Depression Imaging REsearch ConsorTium

(DIRECT) has collected structural and resting-state functional MRI (rsfMRI) data from 2,428 participants, and provided insights into changes in the functional connectivity in patients with depression, though they have not attempted to build predictors of depression status [9].

Another limitation of previous brain imaging studies of depression is that they considered high-level brain measurements, such as cortical thickness and subcortical volumes in regions of interests (ROIs) defined using brain atlases [10]. While this reduces multiple testing and computation, these high-level measurements lack granularity and may fail at detecting subtle and localised associations in the brain [11, 12]. This may explain why approaches that relied on data reduction techniques have failed to yield depression predictors with detectable prediction accuracy [13, 14].

In this project, we aimed to train predictors of major depressive disorders (MDD) on pre-processed high-dimensional (voxel-based) brain MRI data from the UK Biobank [10], using machine-learning and deep-learning algorithms, which can handle high-dimensional data [5, 15]. Although the UK Biobank was not focused on depression, more than a thousand participants with brain MRI data available had a recorded episode of MDD. The large sample and homogeneous image acquisition (compared to the data gathered by consortia) may facilitate the training of algorithms. Our pre-processing includes the normalisation of images into a standard space, which should simplify the learning task and ensure that both deep-learning and machine-learning can be trained on the same images. Lastly, the voxel-based representation of grey matter retains more granularity and signal of interest than a ROI-based approach. To progress in understanding depression-related brain biology, we sought to identify the brain regions contributing to the algorithms' predictive ability. Lastly, we evaluated if the brain-based predicted risk captured independent information from a genetic predictor of depression.

## **Materials and Methods**

An overview of study is presented in Supplementary Figure 1.

### **Code availability**

Code used in this study is available at <https://github.com/baptisteCD/Brain-MRI-based-prediction-of-depression>.

### **Data availability**

Data used in the manuscript are held and distributed by the relevant data managing teams. For the UK Biobank, access is available via <https://www.ukbiobank.ac.uk/use-our-data/apply-for-access/>. Data fields generated in study using the UK Biobank (such as MDD status and processed brain images) may be re-created using the information provided in Methods. The DEP-ARREST CLIN cohort is available upon request, and some data are available as part of the ENIGMA-MDD consortium. The trained BLUP model is available at <https://github.com/baptisteCD/Brain-MRI-based-prediction-of-depression>.

### **UK Biobank**

The UK Biobank cohort consists of over 500,000 volunteers (aged 40-69 years at baseline) from the UK, recruited between 2006 and 2010 [16]. We used the UK Biobank data collected between 2006 and March 2023, and excluded individuals who withdrew from the study. The MRI scan and image acquisition details used by the UK Biobank was previously described in detail [17]. At the time of analysis, the brain images for 39,994 participants were available [18]. After excluding individuals who did not answer or selected “prefer not to answer” for the smoking status question, or lacked information on BMI at baseline, a total of 39,740 individuals were retained for further analysis.

### **MDD status and antidepressant use ascertainment in the UK Biobank**

The full UK Biobank data-fields and entry codes used to define MDD cases are summarised in Supplementary Table 1. Briefly, MDD cases were defined by ICD-10 (data-field 41270) or ICD-9 (data-field 41271) codes for depressive episode, recurrent depressive disorder, persistent mood (affective) disorders, other mood (affective) disorders or unknown mood (affective) disorders. We excluded MDD cases with ICD-10 or ICD-9 codes of bipolar disorder or schizophrenia, had self-reported schizophrenia (data-field 20002), had an ICD-10 code of multiple personality disorder, or were taking antipsychotics (data-field 20003). The screened controls were defined as individuals who had no diagnosis of MDD, bipolar disorder or schizophrenia, had no ICD diagnosis of personality, anxiety or neurotic disorders, had no self-reported history of mood disorders, and were not on any antidepressants or antipsychotic medications. Amongst individuals who had MRI data available and with complete records of age at baseline (calculated using data-field 34 “year of birth” and data-field 53 “Date of attending assessment centre”), sex (data-field 31), BMI (data-field 21001) and smoking status (data-field 20116) at baseline, we identified a total of 1,496 MDD cases and 27,741 screened controls. To evaluate the potential effects of antidepressants on brain structure, we also identified individuals who self-reported to have used antidepressants at the initial or follow-up assessments.

### **Training and test cohorts**

To construct the training cohort, we randomly selected 1,000 MDD cases. Using the “MatchIt” R package (version 4.5.5), we selected four controls for each MDD case, matched by sex, genetic ancestry (ancestry calling is described in Yengo et al. [19]), age at baseline, smoking status (ever smoked or never smoked) at baseline, and BMI at baseline. The test cohort consisted of the remaining MDD cases (N=496) and matched controls not included in the training sample. After

excluding individuals whose brain MRI images were deemed “non-usable” as previously described [18], the training cohort consisted of 987 MDD cases and 3,934 controls, and the test cohort consisted of 483 MDD cases and 1,939 controls (demographic summary is presented in Supplementary Table 2).

### **Pre-processing of brain MRI images**

Using the T1-volume pipeline from the Clinica software platform [20], we processed the T1-weighted MRI images from the UK Biobank with the statistical parametric mapping (SPM) software package [21]. Tissue segmentation, bias correction and spatial normalisation were performed, and the grey matter in the brain images was extracted. A Dartel template was generated from the images of 1,000 randomly selected study subjects from the training cohort, and the brain MRI images of study subjects were normalised and aligned to a common space by projection onto the Dartel template. This step ensured voxel-wise correspondence across the subjects. 2,122,945 voxels per image were included in subsequent analysis, though only a fraction corresponds to the grey matter.

### **Brain atlas for interpretation**

Voxels were mapped to brain anatomical regions using the cerebellum atlas by Diedrichsen et al. [22], and the cortical and subcortical atlases by Makris et al. [23], Frazier et al. [24] and Goldstein et al. [25].

### **Training a machine-learning predictor**

We used the OmicS-data-based Complex trait Analysis (OSCA) tool [26] to train a machine-learning-based predictor of MDD risk using the Best Linear Unbiased Prediction (BLUP) [27]. BLUP is a statistical framework and a special case of machine-learning algorithms that maximises

the restricted maximum likelihood (REML). As such, it does not require cross-validation and is computationally efficient [28].

Voxels with 0 values for most-to-all individuals are on the edge or outside of the grey matter and cannot be fitted in a linear model. Hence, we excluded voxels with mean intensity value  $< 0.1$  and variance  $< 0.01$ , leaving a total of 415,836 voxels from 150 ROIs for the BLUP analysis. The joint, conditional effects of the voxels were estimated from the training cohort using REML, and the voxel weights were applied to the test cohort to generate the BLUP predictor.

In the training cohort, we also estimated the proportion of phenotypic variance captured by all voxel-wise intensities (previously coined “morphometricity” [29]) using OSCA [26]. We tested the statistical significance of a non-zero morphometricity using a likelihood ratio test [26]. The morphometricity model included as covariates age at baseline, age at imaging, sex, genetic ancestry, smoking status and BMI at baseline, as well as a number of MRI imaging covariates, namely mean rfMRI head motion (data-field 25741), scanner lateral (X) brain position (data-field 25756), scanner transverse (Y) brain position (data-field 25757), scanner longitudinal (Z) brain position (data-field 25758), intensity scaling for T1 (data-field 25925), volume of EstimatedTotalIntraCranial (whole brain) (data-field 26521), inverted signal-to-noise ratio in T1 (data-field 25734), and inverted contrast-to-noise ratio in T1 (data-field 25735).

### **Training a deep-learning predictor**

We trained a deep-learning model to perform a classification task that predicts MDD status using the ClinicaDL software platform [30]. Briefly, tensors were extracted from the MRI images to prepare for downstream training that relies on the PyTorch deep-learning library [31]. The training cohort was split into a training and validation set at a 3:1 ratio, while balancing age at baseline, age at imaging, sex, genetic ancestry, smoking status and BMI at baseline, and birth year. We

trained a ResNet3D architecture using a weighted sampler (giving a stronger weight to underrepresented class), 50 epochs and a 10% dropout rate. The model generated in the epoch with the lowest cross-entropy loss on the validation set was selected as the final model for prediction in the test cohort (hereby referred to as the deep-learning predictor).

### **Evaluating prediction accuracy**

We compared the predicted scores derived from the BLUP and deep-learning predictors between the true MDD cases and controls in the test cohort. We also investigated the covariate-adjusted association between the predictors and MDD status using logistic regression, where the predictors were normalised to have a mean of 0 and a standard deviation (SD) of 1. The logistic regression model adjusted for age at baseline, age at imaging, sex (except in sex-stratified analyses), genetic ancestry, and smoking status and BMI at baseline. We reported the odds ratios (OR) from logistic regression. Given the overrepresentation of female cases and controls in our training cohort, we repeated the logistic regression analysis in males and females separately. We used a two-sided p-value  $< 0.0125$  (multiple testing correction for two predictors and two sexes) to declare statistical significance.

We performed sensitivity analysis where the logistic regression model included MRI imaging covariates as additional covariates. As only 61.3% of the MDD cases in the test cohort had self-reported antidepressant use, we repeated the logistic regression adjusting for antidepressant use. We performed additional sensitivity analyses after excluding individuals with a recorded MDD episode (estimated using the date of the ICD-9 and ICD-10 records) more than 5 years prior to or after imaging.

We defined prediction accuracy as the area under the receiver operator characteristic curve (AUC), which can be interpreted as a probability of a case having a higher predicted score than a control.

Using the pROC [32] R package (version 1.18.0), we calculated the AUC of the BLUP and deep-learning predictors (performed with the “roc” function), and compared the AUC of different predictors (e.g. BLUP versus deep-learning) using DeLong's test for two correlated ROC curves (performed with the “roc.test” function) [32].

### **Identifying MDD-associated ROIs from the BLUP predictor**

We built ROI-specific predictors by restricting the score to the voxels in each ROI (using BLUP weights estimated across the whole brain, as in the previous analysis). We used logistic regression to investigate the association between each ROI-specific BLUP predictor and MDD risk, while adjusting for covariates as in the main analysis. A p-value  $< 0.00033$  (Bonferroni correction for 150 ROIs) was used to declare statistical significance.

### **Investigating the association with genetically predicted MD risks**

GWASs have identified many genetic variants implicated in the risk of depression [3, 33, 34]. As an estimate for the lifetime genetic liability of an individual to depression, we developed polygenic scores (PGS) for major depression (MD, different from MDD as MD includes self-reported cases). We used SBayesR [35] with the banded LD matrix reference to derive the per-allele weights from the latest MD GWAS [34], after the exclusion of non-European-ancestry participants as well as participants from UK Biobank and 23andme.

As the SNP weights were estimated from GWAS summary statistics derived from European-ancestry individuals, we calculated the PGS<sub>MD</sub> for the subset of individuals of European ancestry in the UK Biobank test cohort using PLINK1.9 [36].

Using linear regression, we estimated the correlation of PGS<sub>MD</sub> with the BLUP and deep-learning predictors (scaled to a mean of 0 and a SD of 1), adjusting for sex (except in sex-stratified analysis),

age at baseline, age at imaging, smoking and BMI at baseline and 10 genetic principal components (PC). A two-sided p-value  $< 0.0125$  (Bonferroni correction for two predictors and two sexes) was used to declare statistical significance.

### **Replication in the DEP-ARREST CLIN cohort**

We sought to replicate our results in the DEP-ARREST CLIN cohort. The DEP-ARREST CLIN cohort consisted of 64 patients and 32 controls aged 18-65 years old. The cases were hospitalised in the Department of Psychiatry of Bicêtre University Hospital with a diagnosis of current Major Depressive Episode (MDE) in the context of MDD (DSM-IVTR) based on the Mini International Neuropsychiatric Interview (MINI) [37]. Patients with a Hamilton Depression Rating Scale 17 items (HDRS) score  $\geq 18$  were included, and were excluded if they had brain syndromes, unstable medical conditions, bipolar disorders (DSM-IVTR) or current treatment with mood stabilisers, psychotic disorders (DSM-IVTR) or current treatment with antipsychotics, current substance abuse or dependence (DSM-IVTR), as well as pregnancy, breastfeeding, and contra-indications to cerebral MRI. All patients were systematically assessed by a psychiatrist. Information on the MRI scanner and acquisition protocol used in the DEP-ARREST CLIN cohort was previously reported [38].

We processed the brain MRI images using Clinica as per the UK Biobank. Images were normalised to the DARTEL template created from the UK Biobank images. We applied the BLUP weights to the processed grey matter maps to create the predicted scores of MDD risks. We tested the association of the overall and ROI-specific BLUP predictors (that showed at least nominal significance in the UK Biobank) with MDE-MDD status using logistic regression, adjusting for sex, age, BMI, smoking status and intracranial volume (ICV).

### **Results**

### **The BLUP predictor is strongly associated with MDD risk**

As part of the BLUP training, we estimated the morphometricity of the MDD status in the training cohort of UK Biobank, which corresponds to the variance explained ( $R^2$ ) by all grey-matter voxels [29]. The morphometricity was 0.061 (standard error=0.024; p-value= $4.9 \times 10^{-4}$ ), which indicated that the 415,836 voxel-wise intensity measurements accounted for 6.1% of the variance in MDD. This is higher than the morphometricity previously reported for depression score and number of depression symptoms in the UK Biobank cohort [11].

Overall, the BLUP predictor showed an AUC of 0.57 in the test cohort (483 MDD cases and 1,939 controls) (Table 1). Using the BLUP predictor, we found that MDD cases had significantly higher predicted MDD scores compared to the controls (two-sided t-test p-value= $3.1 \times 10^{-5}$ ) (Figure 1A). This difference was confirmed in both females and males (two-sided t-test p-value=0.0028 and p-value=0.0027, respectively) (Figure 1A). The accuracy of the BLUP predictor was confirmed by logistic regression, where after adjusting for covariates, every 1-SD increase in the BLUP predicted risk was significantly associated with a 28% increase in MDD risks (OR=1.28 [1.15-1.43]; p-value= $1.1 \times 10^{-5}$ ) in the mixed-sex cohort (Figure 1A). Similar associations were observed in females (OR=1.23 [1.07-1.41]; p-value=0.0035) and males (OR=1.36 [1.13-1.64]; p-value=0.001) (Figure 1A). We observed comparable OR estimates in sensitivity analyses, when further adjusting for MRI imaging covariates or antidepressant use, and when excluding individuals with a recorded MDD episode more than 5 years prior to or after MRI imaging (Supplementary Figure 2). This indicates that the BLUP predictor is sensitive to active or recent MDD events, and is especially sensitive to recent episodes occurring up to 5 years before imaging (OR=1.43 [1.18-1.73]; p-value=0.00026) (Supplementary Figure 2).

### **BLUP predictor outperforms the deep-learning predictor**

Our ResNet3D deep-learning brain predictor did not show a significant association with MDD risks after multiple testing correction (Figure 1B). Overall, the deep-learning predictor exhibited an AUC of 0.53, indicating a close-to-chance capacity to distinguish MDD cases from controls (Table 1). The BLUP predictor demonstrated better performance than the deep-learning predictor (DeLong's test  $p$ -value=0.037). To test the robustness of the deep-learning training, we trained the model using four other splits of the training and test cohorts, and found that only one split produced a deep-learning predictor with a significant association with MDD status (OR=1.16 [1.05-1.29] in the mixed-sex cohort;  $p$ -value=0.0037) (Supplementary Figure 3).

### **The BLUP predictor identifies MDD-associated brain regions**

To identify ROIs contributing to BLUP prediction, we partitioned the predictor based on brain anatomical regions. At  $p$ -value<0.00033 (multiple testing correction for 150 ROIs), no ROI passed the statistical significance threshold, although 17 ROIs were nominally significantly associated with MDD risks (Figure 2). These included the left and right Crus I (cerebellum), and nine ROIs in the cortex, namely the right parahippocampal gyrus (posterior division), left parahippocampal gyrus (anterior division), left and right Heschls gyri (including H1 and H2), left and right frontal operculum cortices, right frontal medial cortex, as well as left and right central opercular cortices. We also identified six ROIs that corresponded to subcortical structures (right thalamus, left and right hippocampus, left and right amygdala as well as left cerebral white matter [which corresponded to the boundary between the left cerebral white matter and grey matter]) (Supplementary Figure 4).

### **Association with the genetically predicted risk of MDD**

We found that the BLUP predictor showed comparable covariate-adjusted correlation with  $PGS_{MD}$  in the mixed and sex-stratified cohorts (indicated by similar  $\beta$  values), though the correlations only

reached statistical significance in the mixed-sex analysis ( $\beta=0.072$  [0.033-0.11];  $p$ -value=0.0003) and the female cohorts ( $\beta=0.068$  [0.019-0.12];  $p$ -value=0.0071) (Figure 3A). The lack of statistical significance in the male cohort may be due to the reduced sample size. Nevertheless, these results suggested that the genetic and the BLUP predictor captured a common risk component of MDD. In contrast, the deep-learning predicted MDD risk was significantly correlated with PGS<sub>MD</sub> in the female cohort only ( $\beta=0.068$  [0.017-0.12];  $p$ -value=0.0093), but not in the mixed-sex or male cohort (Figure 3B).

Next, we tested if combining PGS<sub>MD</sub> and BLUP scores resulted in additional prediction accuracy over that of a single predictor, where statistical significance was declared at  $p < 0.025$  (multiple testing correction for two additive models: PGS<sub>MD</sub>+BLUP and PGS<sub>MD</sub>+deep-learning). Amongst European-ancestry individuals in the test cohort, the PGS<sub>MD</sub> alone showed an AUC of 0.65. We found that the addition of the BLUP predictor led to a modest, yet significant improvement in predicting MDD risks (AUC<sub>PGS+BLUP</sub>=0.66; DeLong's test for comparison with AUC<sub>PGS</sub>  $p$ -value=0.024) (Table 2). The interaction between PGS<sub>MD</sub> and the BLUP predictor (PGS<sub>MD</sub>\*BLUP) did not improve the AUC above that of the additive model (AUC<sub>PGS\*BLUP</sub>=0.66; DeLong's test for comparison with AUC<sub>PGS+BLUP</sub>  $p$ -value=0.61).

### **Replication in the DEP-ARREST CLIN cohort**

In the independent DEP-ARREST CLIN replication cohort (64 MDE-MDD cases and 32 controls), although the overall BLUP predictor (AUC=0.52) and ROI-specific BLUP predictors were not significantly associated with MDE-MDD status in the DEP-ARREST CLIN cohort, the direction of effects for most ROI-specific BLUP predictors were concordant (i.e. positively associated with MDE-MDD) with the results from the UK Biobank cohort, including the left and right

hippocampus (Supplementary Figure 5). The reduced statistical significance is most likely due to the much smaller sample size and, therefore, much limited statistical power.

## Discussion

The prediction of depression risks remains a challenging task. In this study, leveraging voxel-wise brain MRI imaging data of the grey matter, we developed a BLUP predictor that was significantly associated with MDD status (OR=1.28; p-value= $1.1 \times 10^{-5}$ ), and though the prediction accuracy was modest ( $AUC_{BLUP}=0.57$ ), it held up in both male and female participants. We identified 17 brain regions that were nominally significantly associated with MDD by evaluating BLUP predictors specific to each brain ROI, though none passed multiple testing correction. We found that the BLUP brain predictor contained a genetic component, although it also captured additional signals that are not currently captured by the genetic predictor (supported by nominal significance), which may include brain-mediated environmental effects. In contrast, the deep-learning predictor, developed using a ResNet3D architecture, was only nominally associated with MDD risk and accounted for less variance than the BLUP predictor (OR=1.13;  $AUC_{BLUP}=0.53$ ). We sought to replicate the BLUP results in an independent clinical cohort (DEP-ARREST CLIN). The small sample size and differences in the MRI scans (including scanner brands or acquisition parameters) could potentially contribute to the heterogeneity in findings and prevented any association from reaching significance, although the sign of most ROI-specific associations was concordant with our findings in the UK Biobank.

Previous large-consortium efforts on understanding MDD-related brain biology, such as the ENIGMA consortium [7], have identified a number of brain regions linked to MDD risks. The most consistent findings include a reduction in the hippocampus volume [2]. Furthermore, cell-type enrichment analysis of MDD-related genetic variants identified an enrichment in excitatory

neurons in the midbrain, amygdala, hippocampus, thalamus and cortex [34]. Concordant with these findings, we found that the BLUP predictor specifically derived from the voxel intensities of the hippocampus was nominally significantly associated with MDD risks, further supporting its involvement in MDD aetiology. The hippocampus is a highly plastic region containing neuronal precursor cells in the dentate gyrus that give rise to new granule neurons throughout adulthood [39]. Our ROI-specific BLUP predictors contribute to model interpretability, as they highlight brain regions that most contribute to MDD prediction and could warrant further investigation. The ROI-specific BLUP predictors also implicated brain regions whose implication in MDD remains debated, namely the amygdala and thalamus [2], as well as brain regions whose implications in MDD remain less elucidated, such as Crus I, parahippocampal gyrus, and Heschls gyrus. However, these associations should be interpreted with caution due to the lack of statistical significance after multiple testing correction.

Both genetic and environmental factors play an aetiological role in depression. Previous twin-based studies have estimated a heritability of 35%-40% for MD [40], indicating a moderate contribution of genetic factors. Although its clinical utility remains hindered by the limited prediction performance [41], PGSs, which aggregate the effects of disease-related genetic variants in the genome, have been widely used in research as a measure of lifetime risk of depression [42]. In this study, we showed that the genetically predicted MD risks were correlated with the BLUP brain predictor. We found that the BLUP predictor also captured some MDD risk that the latest  $PGS_{MD}$  (based on hundreds of thousands of cases) [34] does not currently capture, although the improvement in prediction accuracy was rather limited ( $\Delta AUC=0.01$ ). This slight improvement in prediction accuracy could be due to the limited power of the  $PGS_{MD}$  or could reflect the brain-mediated environmental factors that contribute to depression. Several environmental risk factors

have been implicated in increasing MDD risks, such as bullying, substance use, social disadvantage and personal stressful life events (e.g. divorce, financial problems and illness) [43, 44]. These risk factors have been linked to alterations in brain features that are previously implicated in MDD, such as a smaller hippocampal volume, and lower cortical thickness in the insula and the medial orbitofrontal cortex [45, 46, 47, 48]. Many of these regions overlap with or are near the brain ROIs identified in our study to contribute to the BLUP prediction. Further investigation is essential to understand how the morphological changes captured by the BLUP predictor, specifically the changes above and beyond those captured by the genetic predictor, correlate with the environmental aetiology of depression.

In our study, we found that the BLUP predictor demonstrated stronger prediction accuracy for MDD risks compared to the deep-learning predictor. Despite increasing efforts in employing deep-learning approaches in analysing brain MRI images, it was previously reported that with the current sample sizes available, deep-learning methods often performed worse than or comparably as linear or non-linear machine-learning models for prediction at the participant level [49, 50, 51]. This is in contrast with tasks where the prediction is performed at the voxel-level (for instance MRI image segmentation) and for which deep-learning methods are the state of the art, even with relatively modest sample sizes [49, 52]. This is not surprising as voxel-level supervision provides many data points of information per image for training versus only one data point for participant-level supervision.

It should be noted that although significant, the brain-based prediction accuracy of the BLUP predictor remained modest (AUC of 0.57), highlighting the difficulty of classifying MDD status from grey-matter data with the currently available methods, and the need for a large test sample to confirm the prediction statistically and further method improvement to achieve translational

values. Of note, our predictor has demonstrated comparable or exceeding performance than previously reported deep-learning or machine-learning predictors trained on the ENIGMA data [8]. It is important to put in perspective that the prediction accuracy of a BLUP or linear predictor is capped by the morphometricity (which we estimated to be 6.1% of variance in the sample, corresponding to a maximum AUC of 0.64) [53]. This indicates that larger training samples could improve prediction accuracy but may never reach a sufficient accuracy for clinical use. Combining the brain and genetic predictors improved prediction, although it remained limited (AUC=0.66). Small to moderate prediction accuracy can still be useful in research and to progress our understanding of the disease aetiology, such as to study depression risk in cohorts where this information was not available or collected. For proof, thousands of studies in genetics have applied polygenic risk scores, despite a limited prediction accuracy. To further improve the prediction of MDD risk, one may need to consider other brain modalities or measurements, such as the white matter structure or functional connectivity [54, 55], which may be more sensitive to environmental factors, although this may also be a sign of the greater noise in measurements (lower test-retest) [56].

Our study provides novel insights into the development of a grey-matter predictor of MDD risks by applying machine-learning approaches to voxel-based brain MRI imaging data. Previous studies, primarily employing image-derived high-level metrics, are more computationally efficient, though at a cost of granularity. Another novelty of our study is that we demonstrated the novel use of ROI-specific BLUP scores to fine-map brain regions that most contributed to the BLUP predictor, though larger sample sizes are required to statistically confirm the contribution of specific ROIs.

Limitations of our study need to be acknowledged. Firstly, the MDD cases identified in the UK Biobank consisted of a heterogeneous cohort of varying symptoms, treatments, disease severity and onset ages. To address this limitation, we performed sensitivity analyses on the performance of the BLUP predictor, where we additionally adjusted for antidepressant use or restricted the analysis to individuals with a recorded MDD episode within 5 years before or after brain imaging. At the same time, the screened controls in UK Biobank may have substance use disorders, as well as MDD episodes that are not in the records and thereby affect the estimated prediction accuracy. Unlike the UK Biobank, the DEP-ARREST CLIN cohort represents a more clinically homogeneous cohort of MDD patients, as the cases were defined by manual adjudication by psychiatrists and were all medication-free at the time of imaging. Although the small sample size of DEP-ARREST CLIN limited the statistical power to detect any significant association, our findings showed concordant direction of associations between the UK Biobank and DEP-ARREST CLIN cohorts and warrant further replication in larger cohorts with sufficient sample sizes. Secondly, the BLUP predictor was sensitive to recent MDD episodes prior to imaging (indicated by the highest OR), though we also observed a marginal association amongst individuals with MDD episodes up to 5 years after imaging. More work is required to understand whether the brain markers underlying the BLUP predictor represent the causal or consequential effects of MDD episodes and whether the same brain markers contribute to the lifetime risk and future episodes of depression. Thirdly, substantial sex differences have been observed in the prevalence [57], heritability [40], clinical presentations [58] and neuroimaging findings [59] of MDD. Future studies with sufficient sample sizes, where the brain-based predictors are trained on female and male samples separately, are essential to provide insights into the sex-specific aetiology of depression. Lastly, brain regions previously implicated in depression, such as the hippocampus and amygdala, have also been linked

to other psychiatric disorders, such as anxiety and schizophrenia [39, 60]. The overlaps in symptomatology, psychosocial risk factors, neuroimaging features and genetic predisposition make it particularly challenging to distinguish MDD from other psychiatric disorders, such as schizophrenia and bipolar disorder [61]; therefore, the utility of our BLUP predictor and other brain-based predictors in informing differential diagnosis remains an outstanding yet clinically important question, representing a translationally critical avenue for future research.

In conclusion, our study demonstrated the performance of a BLUP predictor in classifying MDD status, and highlighted regions in the brain that potentially contributed to the predictor. Our novel finding of a modest morphometricity underscored the difficulty in training brain-based algorithms that can predict depression status, while suggesting promising avenues for future research, such as exploring other brain imaging modalities and informing differential diagnosis. By integrating genetic information, we showed that our BLUP brain scores likely reflected the joint effects of genetic and environmental risk factors on depression. However, the brain score only yielded a limited improvement in prediction accuracy over that from the genetic predictor. Greater prediction accuracy in distinguishing MDD cases from controls, as well as differentiating MDD from other psychiatric disorders, is required to justify the cost of MRI scans for clinical application. However, similar to the genetic predictors of MDD, our BLUP predictor may serve as a research tool for understanding the underlying predisposition to MDD and its comorbidities. Future studies on the development of symptom-specific predictors will further enrich the current knowledge on the function and pathophysiological links of specific brain regions in MDD. Furthermore, longitudinal studies with medication and treatment information available (such as DEP-ARREST CLIN) are essential for understanding the mechanisms by which pharmacological treatments of depression reverse or improve disease symptoms.

## **Acknowledgements**

Sonia Shah is supported by a Heart Foundation Fellowship. Funding for the DEP-ARREST-CLIN cohort, where Emmanuelle Corruble is the principal investigator, was provided by a national grant (ANR SAMENTA 2012) of the Agence Nationale de la Recherche (ANR) and sponsored by the Institut National de la Santé et de la Recherche Médicale (INSERM) (C13-25, EudraCT 2013-004326-29). The research contributing to these results has received funding from the French government under management of Agence Nationale de la Recherche as part of the “France 2030” program (reference ANR-23-IACL-0008, project PRAIRIE-PSAI), as part of the "Investissements d'avenir" program (reference ANR-19-P3IA-0001, project PRAIRIE 3IA Institute and reference ANR-10-IAIHU-06, project Agence Nationale de la Recherche-10-IA Institut Hospitalo-Universitaire-6), from the European Union’s Horizon Europe Framework Programme (grant number 101136607, project CLARA) and by Inria in the context of the Inria-University of Queensland international team (“Brainetics”). Baptiste Couvy-Duchesne is supported by Inria (starting faculty position) and an NHMRC CJ Martin Fellowship (app 1161356).

## **Participant consent statement and ethics approval**

Informed consent was obtained from all UK Biobank participants by the UK Biobank. Procedures are controlled by a dedicated Ethics and Guidance Council (<http://www.ukbiobank.ac.uk/ethics>), with the Ethics and Governance Framework available at <https://www.ukbiobank.ac.uk/media/0xsbmfmw/egf.pdf>. IRB approval was also obtained from the North West Multi-centre Research Ethics Committee. Our research has been conducted using the UK Biobank Resource under Application Number 12505. This research is covered by The University of Queensland Human Research Ethics Committee approval (HREC number 2020/HE002938).

Written informed consent of the DEP-ARREST CLIN participants was obtained after the nature of the procedures had been fully explained.

### **Conflict of interest**

Competing financial interests related to the present article: none to disclose for all authors.

Competing financial interests unrelated to the present article: Olivier Colliot reports having received consulting fees from Therapanacea. Olivier Colliot reports that other principal investigators affiliated to the team which he co-leads have received grants (paid to the institution) from Sanofi and Biogen. Olivier Colliot reports that his spouse was an employee of myBrainTechnologies and is an employee of DiamPark. Olivier Colliot holds a patent registered at the International Bureau of the World Intellectual Property Organization (PCT/IB2016/0526993, Allasonniere S, Colliot O, Durrleman S, Schiratti J-B, A method for determining the temporal progression of a biological phenomenon and associated methods and devices) (2017). Olivier Colliot has submitted a declaration of invention to the European Patent Office (File Number: EP23306632, Ayache N, Colliot O, Hamzaoui M, Soulier T, Stankoff B, Devices for generation of synthetic 3D representation representative of myelin content, Submitted to the European Patent Office on September 28th, 2023).

### **Author contributions**

JCJ: conceptualisation, methodology, formal analysis, visualisation, writing - original draft, writing - review & editing; CB: Software, writing - review & editing; ED, RC, HB and EC: data curation, writing - review & editing; NRW: supervision, writing - review & editing; OC: supervision, software, writing - review & editing; SS: supervision, writing - review & editing;

BCD: conceptualisation, methodology, validation, data curation, writing - review & editing, supervision, funding acquisition.

ARTICLE IN PRESS

## References

1. Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*. 2020;396(10258):1204-22.
2. Herrman H, Patel V, Kieling C, Berk M, Buchweitz C, Cuijpers P, et al. Time for united action on depression: a Lancet-World Psychiatric Association Commission. *The Lancet*. 2022;399(10328):957-1022.
3. Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci*. 2019;22(3):343-52.
4. Dusi N, Barlati S, Vita A, Brambilla P. Brain structural effects of antidepressant treatment in major depression. *Curr Neuropharmacol*. 2015;13(4):458-65.
5. Patel MJ, Khalaf A, Aizenstein HJ. Studying depression using imaging and machine learning methods. *NeuroImage: Clinical*. 2016;10:115-23.
6. Winter NR, Leenings R, Ernsting J, Sarink K, Fisch L, Emden D, et al. Quantifying deviations of brain structure and function in major depressive disorder across neuroimaging modalities. *JAMA Psychiatry*. 2022;79(9):879-88.
7. Schmaal L, Pozzi E, C. Ho T, van Velzen LS, Veer IM, Opel N, et al. ENIGMA MDD: seven years of global neuroimaging studies of major depression through worldwide data sharing. *Transl Psychiatry*. 2020;10(1):172.
8. Goya-Maldonado R, Erwin-Grabner T, Zeng L-L, Ching CRK, Aleman A, Amod AR, et al. Classification of major depressive disorder using vertex-wise brain sulcal depth, curvature, and thickness with a deep and a shallow learning model. *Mol Psychiatry*. 2025.

9. Chen X, Lu B, Li H-X, Li X-Y, Wang Y-W, Castellanos FX, et al. The DIRECT consortium and the REST-meta-MDD project: towards neuroimaging biomarkers of major depressive disorder. *Psychoradiology*. 2022;2(1):32-42.
10. Harris MA, Cox SR, de Nooij L, Barbu MC, Adams MJ, Shen X, et al. Structural neuroimaging measures and lifetime depression across levels of phenotyping in UK biobank. *Transl Psychiatry*. 2022;12(1):157.
11. Couvy-Duchesne B, Strike LT, Zhang F, Holtz Y, Zheng Z, Kemper KE, et al. A unified framework for association and prediction from vertex-wise grey-matter structure. *Hum Brain Mapp*. 2020;41(14):4062-76.
12. Fürtjes AE, Cole JH, Couvy-Duchesne B, Ritchie SJ. A quantified comparison of cortical atlases on the basis of trait morphometricity. *Cortex*. 2023;158:110-26.
13. Belov V, Erwin-Grabner T, Aghajani M, Aleman A, Amod AR, Basgoze Z, et al. Multi-site benchmark classification of major depressive disorder using machine learning on cortical and subcortical measures. *Sci Rep*. 2024;14(1):1084.
14. Winter NR, Blanke J, Leenings R, Ernsting J, Fisch L, Sarink K, et al. A systematic evaluation of machine learning-based biomarkers for major depressive disorder. *JAMA Psychiatry*. 2024;81(4):386-95.
15. Cao B, Yang E, Wang L, Mo Z, Steffens DC, Zhang H, et al. Brain morphometric features predict depression symptom phenotypes in late-life depression using a deep learning model. *Front Neurosci*. 2023;17:1209906.
16. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-9.

17. Smith SM, Alfaro-Almagro F, Miller KL. UK Biobank Brain Imaging Documentation Version 1.10. Oxford Centre for Functional MRI of the Brain (FMRIB/WIN), Oxford University on behalf of UK Biobank; 2024.
18. Alfaro-Almagro F, Jenkinson M, Bangerter NK, Andersson JLR, Griffanti L, Douaud G, et al. Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage*. 2018;166:400-24.
19. Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. *Hum Mol Genet*. 2018;27(20):3641-9.
20. Routier A, Burgos N, Díaz M, Bacci M, Bottani S, El-Rifai O, et al. Clinica: an open-source software platform for reproducible clinical neuroscience studies. *Front Neuroinform*. 2021;15.
21. Chapter 31 - experimental design and statistical parametric mapping. In: Frackowiak RSJ, Friston KJ, Frith CD, Dolan RJ, Price CJ, Zeki S, et al., editors. *Human Brain Function* (Second Edition). Burlington: Academic Press; 2004. p. 599-632.
22. Diedrichsen J, Balsters JH, Flavell J, Cussans E, Ramnani N. A probabilistic MR atlas of the human cerebellum. *Neuroimage*. 2009;46(1):39-46.
23. Makris N, Goldstein JM, Kennedy D, Hodge SM, Caviness VS, Faraone SV, et al. Decreased volume of left and total anterior insular lobule in schizophrenia. *Schizophr Res*. 2006;83(2):155-71.
24. Frazier JA, Chiu S, Breeze JL, Makris N, Lange N, Kennedy DN, et al. Structural brain magnetic resonance imaging of limbic and thalamic volumes in pediatric bipolar disorder. *Am J Psychiatry*. 2005;162(7):1256-65.

25. Goldstein JM, Seidman LJ, Makris N, Ahern T, O'Brien LM, Caviness VS, et al. Hypothalamic abnormalities in schizophrenia: sex effects and genetic vulnerability. *Biol Psychiatry*. 2007;61(8):935-45.
26. Zhang F, Chen W, Zhu Z, Zhang Q, Nabais MF, Qi T, et al. OSCA: a tool for omic-data-based complex trait analysis. *Genome Biol*. 2019;20(1):107.
27. Henderson CR. Best linear unbiased estimation and prediction under a selection model. *Biometrics*. 1975;31(2):423-47.
28. Couvy-Duchesne B, Faouzi J, Martin B, Thibeau-Sutre E, Wild A, Ansart M, et al. Ensemble learning of Convolutional Neural Network, Support Vector Machine, and Best Linear Unbiased Predictor for brain age prediction: ARAMIS contribution to the predictive analytics competition 2019 challenge. *Front Psychiatry*. 2020;11:593336.
29. Sabuncu MR, Ge T, Holmes AJ, Smoller JW, Buckner RL, Fischl B. Morphometricity as a measure of the neuroanatomical signature of a trait. *Proc Natl Acad Sci U S A*. 2016;113(39):E5749-56.
30. Thibeau-Sutre E, Díaz M, Hassanaly R, Routier A, Dormont D, Colliot O, et al. ClinicaDL: An open-source deep learning software for reproducible neuroimaging processing. *Comput Methods Programs Biomed*. 2022;220:106818.
31. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: an imperative style, high-performance deep learning library. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*: Curran Associates Inc.; 2019. p. Article 721.
32. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12(1):77.

33. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet.* 2018;50(5):668-81.
34. Adams MJ, Streit F, Meng X, Awasthi S, Adey BN, Choi KW, et al. Trans-ancestry genome-wide study of depression identifies 697 associations implicating cell types and pharmacotherapies. *Cell.* 2025.
35. Lloyd-Jones LR, Zeng J, Sidorenko J, Yengo L, Moser G, Kemper KE, et al. Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat Commun.* 2019;10(1):5086.
36. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience.* 2015;4(1):s13742-015-0047-8.
37. Colle R, Segawa T, Chupin M, Tran Dong MNTK, Hardy P, Falissard B, et al. Early life adversity is associated with a smaller hippocampus in male but not female depressed in-patients: a case–control study. *BMC Psychiatry.* 2017;17(1):71.
38. Poirot MG, Boucherie DE, Caan MWA, Goya-Maldonado R, Belov V, Corruble E, et al. Predicting antidepressant treatment response from cortical structure on MRI: a mega-analysis from the ENIGMA-MDD working group. *Hum Brain Mapp.* 2025;46(1):e70053.
39. Schoenfeld TJ, Cameron HA. Adult neurogenesis and mental illness. *Neuropsychopharmacology.* 2015;40(1):113-28.
40. Kendler KS, Gatz M, Gardner CO, Pedersen NL. A Swedish national twin study of lifetime major depression. *Am J Psychiatry.* 2006;163(1):109-14.

41. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med.* 2020;12(1):44.
42. Jiang J-C, Singh K, Nitin R, Davis LK, Wray NR, Shah S. Sex-specific association between genetic risk of psychiatric disorders and cardiovascular diseases. *Circ Genom Precis Med.* 2024:e004685.
43. Lopizzo N, Bocchio Chiavetto L, Cattane N, Plazzotta G, Tarazi FI, Pariante CM, et al. Gene–environment interaction in major depression: focus on experience-dependent biological systems. *Frontiers in Psychiatry.* 2015;6.
44. Colodro-Conde L, Couvy-Duchesne B, Zhu G, Coventry WL, Byrne EM, Gordon S, et al. A direct test of the diathesis–stress model for depression. *Mol Psychiatry.* 2018;23(7):1590-6.
45. Nolfé G, Cirillo M, Iavarone A, Negro A, Garofalo E, Cotena A, et al. Bullying at workplace and brain-imaging correlates. *Journal of Clinical Medicine.* 2018;7(8):200.
46. Crossley NA, Alliende LM, Ossandon T, Castañeda CP, González-Valderrama A, Undurraga J, et al. Imaging social and environmental factors as modulators of brain dysfunction: time to focus on developing non-western societies. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging.* 2019;4(1):8-15.
47. Mackey S, Allgaier N, Chaarani B, Spechler P, Orr C, Bunn J, et al. Mega-Analysis of gray matter volume in substance dependence: general and substance-specific regional effects. *Am J Psychiatry.* 2018;176(2):119-28.
48. Papagni SA, Benetti S, Arulanantham S, McCrory E, McGuire P, Mechelli A. Effects of stressful life events on human brain structure: A longitudinal voxel-based morphometry study. *Stress.* 2011;14(2):227-32.

49. Schulz M-A, Yeo BTT, Vogelstein JT, Mourao-Miranada J, Kather JN, Kording K, et al. Different scaling of linear models and deep learning in UKBiobank brain images versus machine-learning datasets. *Nat Commun.* 2020;11(1):4238.
50. Mihalik A, Brudfors M, Robu M, Ferreira FS, Lin H, Rau A, et al., editors. ABCD Neurocognitive Prediction Challenge 2019: Predicting individual fluid intelligence scores from structural MRI using probabilistic segmentation and kernel ridge regression. *Adolescent Brain Cognitive Development Neurocognitive Prediction; 2019 2019//*; Cham: Springer International Publishing.
51. Wen J, Thibeau-Sutre E, Diaz-Melo M, Samper-González J, Routier A, Bottani S, et al. Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation. *Med Image Anal.* 2020;63:101694.
52. Antonelli M, Reinke A, Bakas S, Farahani K, Kopp-Schneider A, Landman BA, et al. The Medical Segmentation Decathlon. *Nat Commun.* 2022;13(1):4128.
53. A probability-based measure of effect size: robustness to base rates and other factors, (2008).
54. Zhou E, Wang W, Ma S, Xie X, Kang L, Xu S, et al. Prediction of anxious depression using multimodal neuroimaging and machine learning. *Neuroimage.* 2024;285:120499.
55. Li J, Chen H, Fan F, Qiu J, Du L, Xiao J, et al. White-matter functional topology: a neuromarker for classification and prediction in unmedicated depression. *Transl Psychiatry.* 2020;10(1):365.
56. Smith SM, Douaud G, Chen W, Hanayik T, Alfaro-Almagro F, Sharp K, et al. An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nat Neurosci.* 2021;24(5):737-45.

57. Hasin DS, Sarvet AL, Meyers JL, Saha TD, Ruan WJ, Stohl M, et al. Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. *JAMA Psychiatry*. 2018;75(4):336-46.
58. Smith DJ, Kyle S, Forty L, Cooper C, Walters J, Russell E, et al. Differences in depressive symptom profile between males and females. *J Affect Disord*. 2008;108(3):279-84.
59. Mohammadi S, Seyedmirzaei H, Salehi MA, Jahanshahi A, Zakavi SS, Dehghani Firouzabadi F, et al. Brain-based sex differences in depression: a systematic review of neuroimaging studies. *Brain Imaging and Behavior*. 2023;17(5):541-69.
60. Zheng G, Zhou Y, Zhou J, Liang S, Li X, Xu C, et al. Abnormalities of the amygdala in schizophrenia: a real world study. *BMC Psychiatry*. 2023;23(1):615.
61. Stein F, Kircher T. Transdiagnostic findings across major depressive disorder, bipolar disorder and schizophrenia: a qualitative review. *J Affect Disord*. 2025;387:119464.

**Tables**

Table 1. AUC of BLUP and deep-learning predictors in predicting MDD risks in the UK Biobank test cohort

Predictor	AUC	DeLong's test for BLUP versus deep-learning
BLUP	0.57	p-value=0.037
Deep-learning	0.53	

ARTICLE IN PRESS

Table 2. AUC of genetic and brain-based predictors in predicting MDD risks amongst European-ancestry individuals of the UK Biobank test cohort

Predictor	AUC	DeLong's test for comparison with PGS <sub>MD</sub> only
PGS <sub>MD</sub>	0.65	NA
PGS <sub>MD</sub> + BLUP	0.66	0.024
PGS <sub>MD</sub> * BLUP	0.66	0.026
PGS <sub>MD</sub> + deep-learning	0.65	0.42

ARTICLE IN PRESS

## Figure legends

Figure 1. Performance of the BLUP and deep-learning predictors in the UK Biobank test cohort. The association of the (A) BLUP and (B) deep-learning predictors with MDD status was evaluated in the test cohort by (i) comparing the distribution of the predictors between MDD cases and controls using t-tests, and (ii) logistic regression, adjusting for age at baseline, age at imaging, sex (except in sex-stratified analyses), genetic ancestry, smoking status at baseline, and BMI at baseline. OR represents the change in MDD risk per 1-SD increase in the BLUP or deep-learning score. The grey line represents no association (OR=1).

Figure 2. Association between MDD risks and ROI-specific BLUP predictors in the UK Biobank test cohort. Logistic regression was performed to investigate the association between each ROI-specific BLUP predictor and MDD risks. The logistic regression model included as covariates age at baseline, age at imaging, sex, genetic ancestry, smoking status at baseline, and BMI at baseline. OR represents the change in MDD risks per 1-SD increase in the BLUP score. The grey line represents no association (OR=1).

Figure 3. Association of  $PGS_{MD}$  with the BLUP and deep-learning scores in the UK Biobank test cohort. Linear regression was performed amongst individuals of European ancestry in the test cohort to investigate the association of  $PGS_{MD}$  with the (A) BLUP and (B) deep-learning predictors of MDD, adjusting for sex (except in sex-stratified analysis), age at baseline, age at imaging, smoking status at baseline, BMI at baseline and 10 genetic principal components.  $\beta$  represents the covariate-adjusted correlation between the BLUP or deep-learning predictors and  $PGS_{MD}$ .





