Title: Polygenic Risk-Informed White Matter Integrity Improves Deep Learning-Based Prediction of Youth Depression

Jungyoun Janice Min[1], Heehwan Wang[2], Eunji Lee[3], Bo-Gyeom Kim[3], Gakyung Kim[1], Seung Yun Choi[1], Kahyun Choi[1], Sung Hee Hong[4], Yumin Jang[4], Yu Jin Rah[1], Ji Yeon Kim[1], Seunghui Han[5], Kyung Hwa Lee[5], Junhyung Kim[6], Milenna T. van Dijk[7,8], Jae-Won Kim[5], Grace H. Chung[4], Dongil Chung[9], Sang Ah Lee[1], Yoonjung Yoonie Joo[10], Jiook Cha[1,2,3,11] †

Affiliations

1. Department of Brain and Cognitive Sciences, Seoul National University, Seoul, Republic of Korea
2. Interdisciplinary Program in Artificial Intelligence, Seoul National University, Seoul, Republic of Korea
3. Department of Psychology, Seoul National University, Seoul, Republic of Korea
4. Department of Child Development and Family Studies, Seoul National University, Seoul, Republic of Korea
5. Division of Child and Adolescent Psychiatry, Department of Psychiatry, Seoul National University Hospital, Seoul, Republic of Korea
6. Department of psychiatry, Samsung kangbuk Hospital, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
7. Department of Psychiatry, Vagelos College of Physicians and Surgeons, Columbia University, New York, United States of America
8. Division of Translational Epidemiology, New York State Psychiatric Institute, New York, United States of America
9. Department of Biomedical Engineering, Ulsan National Institute of Science and Technology, Ulsan, Republic of Korea
10. Department of Digital Health, Samsung Advanced Institute for Health Sciences & Technology (SAIHST), Sungkyunkwan University, Samsung Medical Center, Seoul, Republic of Korea
11. Institute of Psychological Science, Seoul National University, Seoul, Republic of Korea

Correspondence to:

Jiook Cha, Ph.D.
Gwanak-ro 1, Building #16, Suite M512
Gwanak-gu, Seoul, Republic of Korea
connectome@snu.ac.kr

Keywords: Child Depression, Polygenic Score, White matter, Track-weighted imaging (TWI), Convolutional Neural Network (CNN), Multi cohort validation

Abstract

Early detection of youth depression is crucial, given its rising prevalence and long-term consequences. Although genetic factors contribute significantly to youth depression, their integration with neuroimaging remains limited. We present a deep learning framework using polygenic scores (PGS) to pretrain a 3D convolutional neural network on diffusion MRI (track-weighted fractional anisotropy), capturing gene–brain associations from a multi-ethnic cohort in the Adolescent Brain Cognitive Development Study (N=4,741). Fine-tuned for predicting depression, the model improved cross-sectional (N=266; AUC=0.62) and two-year predictions of depression and suicidality (AUC=0.61–0.66). It outperformed unimodal models, increasing accuracy by up to 24% over PGS-only and 5.8% over brain-only models. Explainable artificial intelligence identified key white matter tracts—superior longitudinal fasciculus, cingulum and corpus callosum—as predictive features. Decision curve analysis showed greater clinical utility. The model generalized to an independent Korean youth sample (N=108; AUC=0.67), supporting the cross-ethnic scalability of PGS-informed diffusion MRI for precision psychiatry

Introduction

Youth depression is a growing global health concern [1-3], marked by early onset, high recurrence, resistance to treatments, and long-term consequences including suicide, which is a leading cause of death of youth [4-9]. These features underscore the urgent need for early identification to alter adverse developmental trajectories. Within the framework of precision psychiatry, which emphasizes individualized risk detection, biological markers reflecting latent vulnerability—such as genetic predispositions and brain-based features—represent a promising avenue for early prediction and intervention.

Youth depression exhibits stronger genetic associations than adult-onset depression [10], including higher polygenic risk scores (PGS) and increased genetic overlap with other psychiatric disorders [11-13]. Twin studies confirm high heritability [14, 15] with genetic factors impacting symptom persistence and stability [10, 16, 17]. Youths with high familial risk are more likely to experience early-onset depression, further shaped by stress and maternal psychopathology [18]. These genetic vulnerabilities interact strongly with neurobiological substrates, particularly white matter integrity, which provides the structural basis for long-range neural connectivity [19-21]. Fractional Anisotropy (FA), derived from diffusion MRI (dMRI), serves as a key marker of white matter integrity [22]. It demonstrates high heritability with genetic factors explaining up to 88% of its variance [19, 23, 24]. Higher polygenic risk for major depressive disorder (MDD) has been linked to reductions in white mater integrity, suggesting that genetic predisposition may contribute to neurostructural vulnerabilities in depression [25]. Specific tracts, such as the corpus callosum, corona radiata, and superior longitudinal fasciculus, are strongly influenced by genetic factors, supporting the notion that white matter differences are under strong genetic control [21, 24].

Converging evidence from human and animal studies implicates disruptions in specific white matter tracts in the pathophysiology of depression. Human studies report reduced FA in major white matter tracts, including the corpus callosum, cingulum, superior longitudinal fasciculus, and uncinate fasciculus [26-32]. Consistent with these findings, animal models of depression show reduced FA in the corpus callosum and related tracts [33, 34]. Developmental alterations in white matter, particularly in the corpus callosum, fimbria, external capsule, and

internal capsule, were observed in rodent model of depression [35]. These findings collectively underscore the critical role of changes in white matter integrity in the pathophysiology of depression.

The interplay between genetic predispositions and white matter integrity is critical for understanding the biological basis of depression. Supporting this, studies have identified significant genetic correlations between measures of diffusion tensor imaging and depression-related traits [23, 36]. Additionally, forebrain white matter-related dMRI measures have been associated with single nucleotide polymorphisms in genes, encoding proteins involved in extracellular matrix and epidermal growth factor signaling, which are critical for myelin repair and synaptic plasticity and closely linked to MDD [20]. These findings suggest shared genetic mechanisms influencing both white matter integrity and depression susceptibility.

Despite advancements, critical gaps remain in our understanding of youth depression. Most studies have examined genetic and neuroimaging factors in isolation, failing to capture their complex interplay [32, 37-39]. Research in other psychiatric disorders shows that integrating genetic and neuroimaging data through multi-modal approaches enhances predictive accuracy over single-modality models [40-42]. Deep learning methodologies enables multi-modal integration [43], by capturing complex non-linear relationships and learning shared representations across different modalities. However, their application in developmental psychiatric research is often constrained by small sample sizes, limiting model generalizability. [44, 45]. Optimized strategies are therefore needed to effectively utilize multi-modal information in youth depression research, where datasets are typically smaller.

Additionally, most existing studies in both neuroimaging and genetics have been conducted on predominantly Caucasian populations, leaving other ethnic groups underrepresented [46, 47]. Addressing this disparity is critical to developing globally applicable predictive models for depression. Further investigation is needed to assess model applicability in underrepresented groups, such as through transfer learning approaches.

To address these challenges, this study introduces a novel deep learning framework that leverages PGS to *pretrain* a model on dMRI data (specifically, track-weighted fractional anisotropy or TW-FA) from a large, multi-ethnic cohort (the Adolescent Brain Cognitive

Development [ABCD] Study), then fine-tune for classification of depression. This approach aims to: (1) improve the accuracy of *both* cross-sectional and longitudinal prediction of youth depression by integrating genetic and neuroimaging information; (2) identify key white matter tracts associated with depression risk; and (3) demonstrate the *cross-ethnic generalizability* of our model by validating it in an independent Korean cohort.

Results

We analyzed 5,248 ABCD participants with multiple ancestral backgrounds meeting strict quality criteria for dMRI and genetic data. The dataset was divided into three groups: Pretrain (N=4,741), Finetune (N=266), and 2-Year Follow-up Prediction (N=236) sets with additional subsets for suicidality analyses (Figure. S1). We pretrained a 3DCNN on depression PGS and composite PGS alongside white matter imaging to learn gene-brain associations using the pretraining set, then fine-tuned for depression classification via 10-fold cross-validation (Figure. 1). Propensity score matching (age, sex, and study site) minimized selection bias in each fold. To ensure robust evaluation of zero-shot prediction in the 2-Year Prediction subsets, we generated 10 matched subsets with diverse controls. Further details on PGS calculations are provided in the Methods section. Full demographics and performance metrics are in Supplementary Materials (Table. S1, S12–S33).

PGS-Pretrained 3DCNN Outperforms Comparison Models in Cross-sectional MDD classification

The PGS-pretrained 3DCNN using track-weighted FA (TW-FA) consistently outperformed unimodal and multimodal baselines for both cross-sectional and 2-year follow-up MDD prediction. For cross-sectional classification, the composite PGS-pretrained model achieved an AUROC of 0.62±0.121 and an F1 score of 0.636±0.072, surpassing models trained with only composite PGS (AUROC = 0.50±0.118, F1 = 0.444±0.149), only TW-FA (AUROC = 0.586±0.133, F1 = 0.278±0.212), and multimodal models trained without pretraining (AUROC = 0.562±0.118, F1 = 0.449±0.217) (Figure. 2A, B). Compared to composite PGS-only models, the pretrained model improved AUROC by 24%, while outperforming brain-only and non-pretrained multimodal models by 5.8% and 10.3%, respectively. F1 score differences were statistically significant across all comparisons ($p < 0.05$) (Table. S13).

The benchmark for psychiatric risk prediction was established using symptom-based model that incorporated behavior (CBCL internalizing problems), age, and biological sex, achieving an AUROC of 0.767±0.07. Our PGS-pretrained model retained 80.5% of this

benchmark performance, demonstrating that genetic and neuroimaging biomarkers can serve as robust predictive features.

A similar pattern was observed with depression PGS, where the pretrained model (AUROC = 0.597±0.12, F1 = 0.629±0.085) outperformed depression PGS-only (AUROC = 0.546±0.114, F1 = 0.517±0.155) and non-pretrained multimodal models (AUROC = 0.54±0.146, F1 = 0.484±0.223) (Figure. S2A, B; Table. S12). Additionally, the pretrained 3DCNN surpassed traditional machine learning models, including support vector classifiers and XGBOOST (Table. S3), further underscoring the effectiveness of integrating PGS-based pretraining with white matter imaging.

PGS-Pretraining Enhances 2-Year Follow-Up MDD Prediction

In the 2-year follow-up MDD prediction, conducted as a zero-shot evaluation on unseen data, our PGS-pretrained model again demonstrated superior predictive accuracy. The 3DCNN+Multi-layer Perceptron(3DCNN+MLP) model, which processes brain images via a composite PGS–pretrained 3DCNN and integrates composite PGS through an additional MLP, achieved an AUROC of 0.61±0.011 and an F1 score of 0.652±0.007, significantly surpassing models using only composite PGS (AUROC = 0.562±0.011, F1 = 0.474±0.016), only TW-FA (AUROC = 0.589±0.012, F1 = 0.534±0.016), and from-scratch multimodal models (AUROC = 0.596±0.008, F1 = 0.604±0.011) ($p < 0.05$) (Figure. 2C, D; Table. S17). Specifically, compared to composite PGS-only models, our pretrained model improved AUROC by 8.5%, while outperforming brain-only and from-scratch multimodal models by 3.6% and 2.3%, respectively.

A similar trend was observed for depression PGS, where the depression PGS-pretrained 3DCNN+MLP model (AUROC = 0.619±0.014, F1 = 0.623±0.012) significantly outperformed depression PGS-only (AUROC = 0.538±0.01, F1 = 0.473±0.017) and from-scratch multimodal models (AUROC = 0.607±0.014, F1 = 0.565±0.02) ($p < 0.05$) (Figure. S2C, D; Table. S16). Importantly, traditional machine learning models showed poor generalizability, with AUROC and F1 scores below 0.55 (Table. S4). In contrast, both PGS-pretrained + MLP models exhibited significant improvements over conventional machine learning models ($p < 0.01$) (Table. S18, S19).

These findings highlight the utility of PGS integration strategies across tasks. While PGS-based pretraining alone improved cross-sectional MDD prediction, likely by embedding stable gene–brain associations, combining pretraining with direct PGS input conferred additional advantages in 2-year follow-up prediction. This suggests that incorporating polygenic risk at both representational and decision levels may be particularly beneficial when forecasting longer-term psychiatric outcomes.

Decision Curve Analysis (DCA) Results for Clinical Utility

To evaluate the potential clinical utility of our pretrained models, we conducted a Decision Curve Analysis (DCA) comparing the default strategies including, “Treat All” and “Treat None” with our best-performing PGS-pretrained 3DCNN. Our model yielded higher net benefit values compared to both comparison strategies across a range of clinically relevant threshold probabilities (0.4-0.5) (Figure. S3 A, B). This indicates superior performance in identifying children who would progress to MDD while minimizing unnecessary interventions.

Leveraging PGS in 2-Year Follow-Up MDD and Suicidality Prediction

The PGS-pretrained 3DCNN demonstrated superior performance in predicting MDD with suicidality at the 2-year follow-up across all tested models. Using composite PGS, the PGS-pretrained 3DCNN + MLP model achieved the best predictive accuracy. For MDD with suicidal active ideation, our model attained an AUROC of 0.64±0.02 and F1 score of 0.675±0.007, showing a 1.9% improvement over composite PGS-only models (AUROC = 0.628±0.015) and a 13.9% improvement over unimodal TW-FA models (AUROC = 0.562±0.021) (Figure. 2E; Table. S5). While symptom-based models reached an AUROC of 0.815±0.021, our model retained 79% of this benchmark, highlighting its utility even in the absence of behavioral indicators. For MDD with suicidal attempts, the PGS-pretrained model achieved an AUROC of 0.665±0.071 (F1 = 0.702±0.034), improving upon from-scratch 3DCNN + MLP models by 9% (AUROC = 0.611±0.061) (Figure. 2F; Table. S7). Compared to the symptom-based model (AUROC =

0.756±0.03), our model retained 88% of its predictive accuracy, further supporting its clinical relevance.

A similar pattern was observed with depression PGS, where the best-performing model using TW-FA and depression PGS reached an AUROC of 0.651±0.021 (F1 = 0.664±0.008), exceeding depression PGS-only (AUROC = 0.614±0.016) by 6% and TW-FA-only models (AUROC = 0.562±0.021) by 15.8% (Figure. S2 E, F; Table. S5). For MDD with suicidal attempts, the depression PGS-pretrained model (AUROC = 0.67±0.076, F1 = 0.651±0.034) outperformed from-scratch multimodal models (AUROC = 0.661±0.074) by 7.4% and TW-FA-only models (AUROC = 0.573±0.062) by 16.9% (Figure. S2 G, H; Table. S7). Across all comparisons, both PGS-pretrained models significantly outperformed alternative approaches in F1 scores ($p < 0.01$) (Table. S20, S21).

Traditional machine learning models failed to generalize in the unseen dataset for MDD and suicidality prediction, with AUROC and F1 scores below 0.55 (Table. S6, S8). In contrast, PGS-pretrained + MLP models exhibited statistically significant improvements over conventional machine learning models ($p < 0.05$) (Table. S22, S23, S26, S27). Notably, our approach maintained robust performance across all comparison models in predicting MDD onset at the 2-year follow-up, even when excluding cases with suicidality.

We conducted a control analysis to evaluate the specificity of PGS-based pretraining. We compared our approach against alternative pretraining approaches using family history of depression, cognitive variable (NIH Toolbox), and behavioral measure (CBCL Total Problems) as pretraining outcomes. The composite PGS-pretrained model showed significantly higher AUROC than the family history and cognitive variable models ($p<0.05$) (Table. S28), while no statistically significant difference was observed compared to the behavioral variable-based pretraining. Additional control analyses, including microstructural versus macrostructural feature evaluations, MDD subgroup analyses and performance assessments of multi-modal models using TW-FA, PGS, age, sex, and behavioral variables are provided in the Supplementary Materials (Figure. S4-S6; Table. S9-S11).

Machine Learning Interpretation

Using Integrated Gradients [48] and SmoothGrad [49], we generated voxel-level saliency maps to identify key white matter features associated with MDD (Figure. 3A). Lower fractional anisotropy (FA) across widespread white matter regions is associated with a higher probability of MDD. To improve interpretability, we performed fasciculus-level analyses, revealing that the superior longitudinal fasciculus, cingulum, corticospinal tract, and corpus callosum contained a high proportion of voxels with reduced FA, significantly differing from controls (Figure. 3B). In contrast, relatively few voxels showed a positive association with MDD (Figure. 3C). Children who developed depressive symptoms and suicidal ideation two years later exhibited similar FA reductions in these tracts, which strongly contributed to their symptom prediction (Figure. 3D–F).

Generalizability testing on Independent Korean Cohorts

To assess cross-ethnic generalizability, we evaluated the PGS-pretrained 3DCNN on an independent Korean adolescent sample. Fine-tuning the PGS-pretrained 3DCNN models on this dataset significantly outperformed the from-scratch model trained solely on Korean data in AUROC (p-value < 0.05) (Table. S33). The composite PGS-pretrained model achieved an AUROC of 0.673 ± 0.154 and F1 score of 0.61 ± 0.227, compared to the from-scratch model's AUROC of 0.532 ± 0.127 and F1 score of 0.5 ± 0.215 (Figure. 2G, H; Table. 3). These observations underscore the value of combining PGS with neuroimaging data to improve cross-population applicability.

Discussion

This study demonstrates that integrating PGS with white matter neuroimaging significantly enhances the prediction of youth depression. Using a deep learning framework, we achieved robust predictive performance in cross-sectional and longitudinal datasets. PGS-pretrained models significantly outperformed alternative pretraining outcomes (e.g. family history, cognitive variables), underscoring the specificity. Our model successfully identified key white matter tracts, including the superior longitudinal fasciculus, cortico-spinal tract, corpus callosum, and cingulum, as critical predictors of depression risk. Furthermore, the model demonstrated strong generalizability, accurately classifying children and adolescents at risk for depression in the independent Korean cohort. To our knowledge, this is the first study to pretrain white matter representations using polygenic risk and apply them to psychiatric outcome prediction, with validation conducted longitudinally and cross-cohort.

Pretraining the deep learning model with PGS likely enhanced the model's capacity to detect biologically meaningful white matter features shaped by genetic risk for depression. Compared to auxiliary pretraining targets such as family history of depression or cognitive variables, our approach yielded the higher classification performance, underscoring the specificity of genetic information in guiding the white matter representation learning.

Depression related PGS captures the cumulative effects of thousands of common variants associated with depression risk, many of which influence neurodevelopmental processes relevant to white matter, including myelination, axonal integrity, and synaptic signaling. Variants such as 5-HTTLPR [50, 51], BDNF [50, 52], SLC6A15 [53], FKBP5-T [54], and COMT [55] have been associated with alterations in white matter microstructure. Beyond individual variants, PGS for depression correlates with FA values, supporting broader genetic influences beyond single-gene effects on white matter integrity [27]. Epigenetic modifications in genes like SLC6A4, COMT, and BDNF [33, 93] may further shape neural pathways by regulating gene expression related to myelination and neural plasticity. During pretraining, the model may have learned to recognize subtle, spatially distributed FA patterns shaped by these genetic factors, enabling more sensitive detection of depression-related white matter deviations during fine-tuning. This approach bridges

genetic vulnerability and neurodevelopmental architecture, offering a biologically grounded framework for early risk identification.

Our approach not only yielded improved predictions of depression but also more nuanced insights about neurological correlations of child depression. We observed a widespread distribution of voxels with reduced FA that were significant predictors of youth depression. Individuals with early-onset MDD, particularly those with recurrent or persistent symptoms, showed widespread reductions in FA, which are also broadly influenced by genetic factors [19, 51, 56, 57]. These results emphasize that whole-brain FA alterations may mediate genetic influences on the development of depression.

Despite the importance of whole-brain level white matter microstructures in MDD prediction, we specified several key fasciculi with a high proportion of reduced FA that were critical for predicting youth depression. The superior longitudinal fasciculus, with over 90% of its voxels exhibiting lower FA in youth with an increased likelihood of MDD compared to controls. As a tract connecting frontal, parietal, temporal, and occipital lobes, the superior longitudinal fasciculus has been consistently linked to depression severity, even among asymptomatic youth with familial risk [39, 58]. Meta-analyses further observed the association between the superior longitudinal fasciculus integrity with depression duration and treatment response [59]. Moreover, reduced superior longitudinal fasciculus integrity has been tied to cognitive symptoms related to depression, including deficits in working memory, executive functioning, and attention maintenance [60, 61]. Consistent with prior findings, our results identify the superior longitudinal fasciculus as the most important tract for predicting depression, both cross-sectionally and longitudinally, underscoring its pivotal role in the neurobiology of youth depression.

The cingulum was another significant predictor, with a high proportion of lower FA values contributing critically to the MDD classification. This tract connects the anterior cingulate cortex to the hippocampus and is central to emotional regulation, self-initiative processes, and spatial memory [62, 63], function closely tied to depressive symptoms such as anhedonia, rumination and memory deficits. Disrupted connectivity in the anterior cingulate cortex is a hallmark of adolescent depression [64]. Furthermore, reduced FA in the cingulum has also been

consistently associated with suicidal thoughts and behaviors, implicating fronto-limbic dysfunction in both mood dysregulation and self-harm risks [65, 66].

The corpus callosum is a well-established white matter marker in depression, with reduced FA frequently linked to the future symptom severity [32, 67, 68]. Among its subregions, the genu and body show the most pronounced reductions in FA, reflecting impaired interhemispheric connectivity in the frontal lobes, which is critical for emotional and cognitive regulation [69-71]. Animal studies further support this, with reduced FA observed in mice lacking the serotonin transporter gene SLC6A4, which is strongly associated with depressive behaviors [34, 72]. These findings, corroborated by human studies [73], emphasize the central role of the corpus callosum in depression-related white matter alterations.

While our model's AUROCs in the 0.61-0.66 range fall within the moderate performance band, they consistently surpassed unimodal and conventional machine learning baselines. In the context of low-prevalence, high-stakes conditions such as youth depression [7], even modest improvements in future outcome prediction can provides incremental predictive value. Decision curve analysis (DCA) [74] further demonstrated that our models offer higher net benefit over default strategies such as "treat-all" or "treat-none" approaches. These findings support the clinical utility of biologically informed prediction models, not as stand-alone diagnostics, but as components of multi-tiered screening frameworks.

Our core thesis is that polygenic risk–informed white matter analyses represent a pivotal step forward in the early detection of youth depression, surpassing conventional unimodal approaches in both effect size and cross-ethnic scalability. To aid interpretation, we converted AUROC values to Cohen's d using the formula $AUROC = \phi\frac{\mathrm{d}}{\sqrt{2}}$ (under the equal variance assumption) [75]. We observed estimated cohen's D of 0.43 (cross-sectional), 0.40 (2-year MDD), 0.51 (suicidal ideation), 0.60 (suicidal attempt), and 0.63 (Korean cohort). These values exceed those reported in large-scale studies where univariate brain features alone often yield modest or non-replicating effects. For example, Van Velzen et al. (2020) reported subtle white matter alterations in adult MDD with d = 0.12–0.26, but adolescent findings did not survive correction [29]. Similarly, large-scale ENIGMA meta-analysis have reported cortical surface area reductions in adolescents with MDD, with effect sizes ranging from $d = -0.26$ to -0.57, as well

as subcortical volume differences, including $d = -0.20$ for the hippocampus [76, 77]. In contrast, our gene-informed white matter framework yields medium-to-large effect sizes across cohorts and outcomes, comparable to or exceeding those reported in prior large-scale studies. This supports its potential as a robust tool for identifying youth at risk of depression.

Our moderate-to-large effect sizes persist even in the smaller independent Korean cohort (N=108), highlighting how transfer learning can bolster predictive power in underrepresented samples. Taken together, these findings underscore the promise of gene-informed white matter models for identifying youth at risk of depression – offering a more robust and generalizable signal than has typically been observed with single-modality or univariate approaches.

While dMRI is a relatively costly modality for large-scale screening, its use may be justified in targeted settings—particularly for youth already flagged by clinical or genomic indicators. By combining genetic predispositions and neurobiological markers, our framework demonstrates the feasibility of early, biology-informed identification. Future work incorporating behavioral assessments and environmental exposures may further improve individualized risk prediction and inform early intervention strategies for high-risk youth.

While this study provides valuable insights into youth depression, several limitations remain. First, our framework does not incorporate environmental factors like early stress and socio-economic status. Future work should incorporate these to explore gene-environmental interactions. Second, dMRI has known limitations in resolving crossing fibers [78], although we applied recommended preprocessing pipelines and utilized track-weighted imaging, which has demonstrated higher sensitivity compared to conventional diffusion tensor imaging in representing crossing fiber configurations [79, 80], to mitigate this. Third, our approach excluded other biological modalities, such as brain hippocampal volume and functional connectivity, limiting a comprehensive view of depression. Future research should address these limitations by integrating environmental measures and diverse imaging modalities to enhance prediction and early intervention strategies.

Despite these limitations, our multi-modal integration strategy advances understanding of pediatric depression by demonstrating how PGS-based pretraining can improve neuroimaging-based classification and prognosis in large-scale cohorts. By integrating genetic predispositions

312 and white matter neuroimaging data, we provide a novel framework for identifying children with
313 depression and further at-risk youth before symptom onset. Our findings highlight multiple white
314 matter tracts—including the corpus callosum, superior longitudinal fasciculus, cingulum, and
315 corticospinal tract—as key features for depression prediction. Furthermore, the model's robust
316 performance in an independent Korean population emphasizes its potential for cross-population
317 generalizability. These findings support a precision psychiatry approach, integrating genetics and
318 neuroimaging for early depression risk detection.

319

Methods

Participants

We used participants from the Adolescent Brain and Cognitive Development (ABCD) study [81] (N=11,868), and a Korean cohort (N=180) recruited through research collaboration with Seoul National University and Ulsan National Institute of Science and Technology. Following rigorous quality control and exclusion criteria, the final analysis included 5,248 ABCD participants and 108 Korean participants. Detailed demographics and full inclusion/exclusion criteria are provided in Supplementary Materials (Table. S1).

Track-weighted FA (TW-FA) Imaging Processing

Track-weighted FA (TW-FA) was used as the input in all experiments. dMRI provides measures of brain macrostructure (e.g. volume(mm^3), length(mm), area(mm^2)) and microstructure (e.g. white matter integrity), non-invasively [82]. The processing was conducted using software based on the MRtrix3 package [83] especially using the command *tckmap* after general MRI quality assessment. TW-FA maps were formed as described in previous studies [79, 80], so we will describe them briefly here. In summary, reference maps (here, FA maps) values were computed at each point along the streamline. For each voxel, a Gaussian-weighted mean of the reference map values was determined within the neighborhood along individual fiber streamlines (yielding the track-wise statistic). The voxel intensity for the track-weighted imaging map was then calculated as the average of these track-wise values across all tracts passing through the voxel, resulting in the voxel-wise statistic. We utilized full-width-half-maximums (FHWM) of 15mm for gaussian neighborhood weightings, which is known to be more sensitive to focal abnormalities [84]. The TW-FA framework applies directional smoothing to tractograms by assigning each streamline a weight derived from the average FA value along its path, effectively creating a track-informed version of the FA map [84, 85]. TW-FA processing was conducted using the same protocol for both cohorts.

348 **Demographic and Behavioral Variables**

349 *Psychiatric Variables of ABCD Cohort*

350 For assessing the children's lifetime MDD and depressive symptoms, our analysis employed
351 the computerized version of the Kiddie-Structured Assessment for Affective Disorders and
352 Schizophrenia (KSADS-COMP) based on child and parent reports and Child Behavioral
353 Checklist (CBCL) [86], which relies on parent reports. KSADS-COMP has been well validated
354 from previous study showing good to high reliability (AUC 0.89-1.0) compared against clinician
355 administered [87]. For the current analysis, we used a lifetime MDD diagnosis variable
356 encompassing past, present and remitted MDD diagnoses, based on summary ratings from the
357 KSADS-COMP, which considers both child and parent reports. We excluded participants who
358 were taking anti-depressant medication and who met lifetime psychotic (e.g. schizophrenia) or
359 neurodevelopmental disorder (e.g. autism spectrum disorder) diagnoses [88, 89]. For a control
360 group, we included participants who met none of the diagnosis criteria in child nor parent report
361 past or current time point and met healthy control criteria of CBCL (total problems T-score<60)
362 from the baseline to 2-year follow-up. We utilized lifetime indicators of suicidality (including
363 suicidal ideation and suicide attempts) based on KSADS-COMP diagnoses, derived from child
364 and parent reports. The criteria for the control group were consistent with those described earlier.

365 Since we excluded all the MDD diagnosed participants from pretraining set to maximize the
366 sample size of the finetune set, depressive symptoms score was used exclusively within the
367 pretraining set to estimate the composite polygenic scores. We generated a depressive symptoms
368 score based on a previous study [90], using *module 1* of the KSADS-COMP symptoms
369 assessment, based on summation of the child and parent reports. Detailed methods for composite
370 PGS calculations are described in the Methods section of *Genetic Data Preprocessing*.

371 In conclusion, 4741 participants were included in the pretraining dataset, and 266
372 participants (control:133) whose propensity scores were matched for age, biological sex, and
373 study site were used for cross-sectional MDD classification in fine-tune set. In addition, to
374 statistically evaluate the 2-year follow-up zero-shot prediction performance, we utilized
375 individuals with MDD diagnosis at the 2-year follow-up and matched them with randomly
376 sampled from participants not included in either the pretrained or finetune sets (control: 118).

This matching process was repeated 10 times to ensure robustness through different samples, which leads to 10 datasets of 236 participants. For suicidality we used the same strategy, resulting in 10 datasets of 116 (control: 58) participants with MDD and suicidal active ideation prediction at the 2-year follow-up and 30 (control: 15) participants with MDD and suicidal attempt prediction. However, we ensured that none of the participants overlapped with pretraining nor finetune sets. Detailed demographics statistics are provided in Supplementary Materials (Table. S1).

Psychiatric Variables of Korean Cohort

Validated Korean translations of depression measures were used to classify participants [91]. Depression was assessed using the child-reported Center for Epidemiological Studies-Depression Scale (CES-D) [92, 93] and parent-reported CBCL [94]. Based on prior meta-analyses, a CES-D cut-off of ≥21, which accounts for gender and racial variations [95, 96], was chosen to align with ABCD study criteria. The depressed group included individuals with CES-D ≥21, while the control group had CES-D <16 and CBCL total problems T-scores <60. The final sample consisted of 108 participants (41 depressed, 67 control). For deep learning, data were split into 10-folds using iterative stratification by age, sex, and CES-D score, followed by propensity score matching for validation and test sets.

Genetic Data Preprocessing

In this study, we used polygenic scores (PGS) to estimate individual-specific genetic propensities, which were previously calculated and validated in [97], where a detailed description of the methodology is provided. For further details on genotype preprocessing, see Supplementary Materials. Here, we briefly summarize the key aspects of the PGS calculation. The PGS were calculated using PLINK version 1.9 and the parameters were optimized in PRS-CSx with Bayesian approach [98]. Since our study focused on youth depression in the American population of multi-ancestry, we utilized PGS scores derived from publicly available multi-ancestry Genome-Wide Association Study (GWAS) summary statistics with robust statistical

power. PGS related to depression, body mass index, post-traumatic stress disorder, alcohol dependency and schizophrenia were included, as these were constructed using multi-ancestry GWAS summary statistics to ensure applicability across diverse populations.

We included composite PGS in our analysis, guided by prior research, that demonstrated the increased predictive power of composite PGS in enhancing risk stratification [99, 100]. To address multicollinearity, we employed Elastic Net, a robust regularization technique, to estimate beta coefficients for each PGS using KSADS-COMP symptom scores as the outcome variable. The model was fitted to maximize explained variance (R^2) while minimizing the number of independent variables through a stepwise elimination process across 5,000 resampled train-validation-test splits. The final composite score was then computed by summing the selected PGS, each weighted by its corresponding beta coefficient from the best-performing model. We ensured that none of the participants in pretraining sets overlapped with finetune or prediction sets.

Model Optimization and Evaluation

Deep Neural Network Training

We implemented a 3D-Convolution Neural Network (3DCNN) model with densenet3D121 as the backbone architecture across all our experiments. The model consists of an initial 3D convolutional module, four 3D dense-blocks, three 3D Transition layers and one fully connected classifier for each task. In the field of medical imaging, prior studies have demonstrated the superior performance of densenet3D121 model compared to traditional CNNs in tasks such as brain tumor classification [101], and Alzheimer's disease diagnosis [102]. Additionally, we conducted experiments using a model that incorporated both each participant's PGS and brain imaging data as input features. This was done by adding the PGS before passing through the fully connected layer of the 3DCNN model, followed by processing through a MLP (3DCNN+MLP). We optimized the model with Mean Squared Error loss for regression tasks and Cross Entropy for the classification tasks. Minimal augmentation was applied, including resizing images to 128x128x128, Random Rotation by 90 degrees (probability=0.5), Random Axis Flip (probability=0.5), and Random Affine Transform (probability=0.5). The deep neural networks

were optimized using the AdamW optimizer with a cosine learning rate decay scheduler and a linear warm-up. Additionally, we implemented early stopping with a patience window of 30 epochs to prevent overfitting on the training data.

We employed two types of comparison models to evaluate our approach. The first set of comparison models focused on modality, including unimodal approaches that used either genetic or neuroimaging data alone and a multi-modal model that incorporated both but without pretraining. This comparison aimed to assess whether both genetic and neuroimaging information are essential for detecting children with MDD and to evaluate the benefit of PGS-based pretraining. Specifically, we implemented an MLP model using only PGS as input, a 3DCNN model using only TW-FA, and a from-scratch 3DCNN+MLP model that integrated both PGS and TW-FA as inputs without employing a pretraining strategy. The second set of comparison models included traditional machine learning models: support vector machine with a linear kernel (SVC_linear), support vector machine with an RBF kernel (SVC_rbf), and eXtreme Gradient Boosting (XGBOOST) using FA-weighted structural connectivity and PGS as their inputs. All experiments were conducted using 10-fold cross-validation, ensuring that the same training, validation, and test folds were used consistently across all experiments. Details on the training parameters for pretraining, cross-sectional and 2-year follow-up MDD prediction, along with transfer learning, are available in the Supplementary Materials.

Explainable Artificial Intelligence

To identify brain regions contributing most to the model's predictions, we employed explainable AI (XAI) techniques, specifically Integrated Gradients [48] combined with SmoothGrad [49]. Integrated Gradients generate saliency maps by integrating the gradients of the model's predictions with respect to the input features, providing a measure of each voxel's contribution. SmoothGrad enhances this approach by adding Gaussian noise ($\sigma = 0.05$) to input images within brain regions, generating 50 noise-augmented images per participant and averaging their resulting saliency maps. This method was implemented using the Captum library (version 0.6.0).

We performed tract-level analyses and identified key tracts containing voxels that significantly contributed to depression classification and prediction. We analyzed only the voxels

with statistically significant XAI attribute values between the control and depressed groups. We conducted voxel-wise t-tests with False Discovery Rate (FDR) corrections using the Benjamin-Hochberg method and applied a cluster-based threshold to identify statistically significant regions. Then we calculated the proportion of significant voxels for each tract, by dividing the total number of significant voxels by the total voxels in the tract. This allowed us to determine the positive and negative associations of each tract in the depression classification task.

Decision Curve Analysis

To assess the clinical utility of our model, we conducted Decision Curve Analysis (DCA) by computing net benefit across a range of clinically relevant threshold probabilities [74]. Net benefit quantifies the balance between true positives and false positives, incorporating the relative harm of unnecessary interventions. It is calculated as:

$$\text{Net Benefit} = \frac{\text{True Positives} - \text{False Positives} \times \frac{\mathrm{p_t}}{1 - \mathrm{p_t}}}{\mathrm{N}}$$

where p is the threshold probability, and N is the sample size. We compared our best-performing models against two baseline strategies: 'Treat All' (assuming all individuals are at risk and should receive intervention) and 'Treat None' (assuming no individuals require intervention). A higher net benefit across threshold probabilities indicates superior clinical applicability by improving the identification of at-risk individuals while minimizing unnecessary interventions.

Acknowledgements

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. 2021R1C1C1006503, RS-2023-00266787, RS-2023-00265406, RS-2024-00421268, RS-2024-00342301), by Creative-Pioneering Researchers Program through Seoul National University(No. 200-20240057, 200-20240135), by Semi-Supervised Learning Research Grant by SAMSUNG(No.A0342-20220009), by Identify the network of brain preparation steps for concentration Research Grant by LooxidLabs(No.339-20230001), by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) [NO.RS-2021-II211343, Artificial Intelligence Graduate School Program (Seoul National University)] by the MSIT(Ministry of Science, ICT), Korea, under the Global Research Support Program in the Digital Field program(RS-2024-00421268) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation), by the National Supercomputing Center with supercomputing resources including technical support(KSC-2023-CRE-0568) and by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A3A2A02090597, NRF-2023R1A2C2005587, RS-2023-00301976), Seoul National University Center for Happiness Studies (No. 0404-20220001), SNU Creative-Pioneering Researchers Program (339-20230014), Research Grant from Seoul National University(350-20230080) and by Artificial intelligence industrial convergence cluster development project funded by the Ministry of Science and ICT(MSIT, Korea) & Gwangju Metropolitan City.

Conflict of Interest

The authors have no conflicts of interest to disclose.

References

1. Mangione, C.M., et al., *Screening for depression and suicide risk in children and adolescents: US preventive services task force recommendation statement.* Jama, 2022. **328**(15): p. 1534-1542.
2. Patton, G.C., et al., *Our future: a Lancet commission on adolescent health and wellbeing.* The Lancet, 2016. **387**(10036): p. 2423-2478.
3. Collaborators, G.M.D., *Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019.* The Lancet Psychiatry, 2022. **9**(2): p. 137-150.
4. Parker, G., et al., *Distinguishing early and late onset non-melancholic unipolar depression.* Journal of Affective Disorders, 2003. **74**(2): p. 131-138.
5. Fleisher, W.P. and L.Y. Katz, *Early onset major depressive disorder.* Paediatrics & Child Health, 2001. **6**(7): p. 444-448.
6. Organization, W.H., *Depression and other common mental disorders: global health estimates.* 2017.
7. Copeland, W.E., et al., *Associations of childhood and adolescent depression with adult psychiatric and functional outcomes.* Journal of the American Academy of Child & Adolescent Psychiatry, 2021. **60**(5): p. 604-611.
8. Toenders, Y.J., et al., *Neuroimaging predictors of onset and course of depression in childhood and adolescence: A systematic review of longitudinal studies.* Developmental cognitive neuroscience, 2019. **39**: p. 100700.
9. Prevention, C.f.D.C.a. *Health Disparities in Suicide*. 2025 [cited 2025 March 23]; Available from: https://www.cdc.gov/suicide/disparities/index.html.
10. Harder, A., et al., *Genetics of age-at-onset in major depression.* Translational Psychiatry, 2022. **12**(1): p. 124.
11. Kendler, K.S., et al., *Age at onset and familial risk for major depression in a Swedish national twin sample.* Psychological Medicine, 2005. **35**(11): p. 1573-1579.
12. Power, R.A., et al., *Genome-wide association for major depression through age at onset stratification: major depressive disorder working group of the psychiatric genomics consortium.* Biological psychiatry, 2017. **81**(4): p. 325-335.
13. Rice, F., et al., *Characterizing developmental trajectories and the role of neuropsychiatric genetic risk variants in early-onset depression.* JAMA psychiatry, 2019. **76**(3): p. 306-313.
14. Thapar, A. and P. McGuffin, *A twin study of depressive symptoms in childhood.* The British Journal of Psychiatry, 1994. **165**(2): p. 259-265.
15. Nivard, M., et al., *Stability in symptoms of anxiety and depression as a function of genotype and environment: a longitudinal twin study from ages 3 to 63 years.* Psychological medicine, 2015. **45**(5): p. 1039-1049.
16. Franić, S., et al., *Childhood and adolescent anxiety and depression: beyond heritability.* Journal of the American Academy of Child & Adolescent Psychiatry, 2010. **49**(8): p. 820-829.

17. Akingbuwa, W.A., et al., *Genetic associations between childhood psychopathology and adult depression and associated traits in 42 998 individuals: a meta-analysis.* JAMA psychiatry, 2020. **77**(7): p. 715-728.
18. Elsayed, N.M., et al., *The role of familial risk, parental psychopathology, and stress for first-onset depression during adolescence.* Journal of affective disorders, 2019. **253**: p. 232-239.
19. Kochunov, P., et al., *Genetics of microstructure of cerebral white matter using diffusion tensor imaging.* Neuroimage, 2010. **53**(3): p. 1109-1116.
20. Elliott, L.T., et al., *Genome-wide association studies of brain imaging phenotypes in UK Biobank.* Nature, 2018. **562**(7726): p. 210-216.
21. Zhao, B., et al., *Common genetic variation influencing human white matter microstructure.* Science, 2021. **372**(6548): p. eabf3736.
22. Johansen-Berg, H. and T.E. Behrens, *Diffusion MRI: from quantitative measurement to in vivo neuroanatomy*. 2013: Academic Press.
23. Zhao, B., et al., *Large-scale GWAS reveals genetic architecture of brain white matter microstructure and genetic overlap with cognitive and mental health traits (n= 17,706).* Molecular psychiatry, 2021. **26**(8): p. 3943-3955.
24. Kochunov, P., et al., *Heritability of fractional anisotropy in human white matter: a comparison of Human Connectome Project and ENIGMA-DTI data.* Neuroimage, 2015. **111**: p. 300-311.
25. Whalley, H.C., et al., *Polygenic risk and white matter integrity in individuals at high risk of mood disorder.* Biological psychiatry, 2013. **74**(4): p. 280-286.
26. Chen, G., et al., *Disorganization of white matter architecture in major depressive disorder: a meta-analysis of diffusion tensor imaging with tract-based spatial statistics.* Scientific reports, 2016. **6**(1): p. 21825.
27. Bracht, T., D. Linden, and P. Keedwell, *A review of white matter microstructure alterations of pathways of the reward circuit in depression.* Journal of affective disorders, 2015. **187**: p. 45-53.
28. Bhatia, K.D., et al., *Reduced integrity of the uncinate fasciculus and cingulum in depression: a stem-by-stem analysis.* Journal of Affective Disorders, 2018. **235**: p. 220-228.
29. Van Velzen, L.S., et al., *White matter disturbances in major depressive disorder: a coordinated analysis across 20 international cohorts in the ENIGMA MDD working group.* Molecular psychiatry, 2020. **25**(7): p. 1511-1525.
30. Shen, X., et al., *Subcortical volume and white matter integrity abnormalities in major depressive disorder: findings from UK Biobank imaging data.* Scientific reports, 2017. **7**(1): p. 5547.
31. Uchida, M., et al., *Association between frontal cortico-limbic white-matter microstructure and risk for pediatric depression.* Psychiatry Research: Neuroimaging, 2021. **318**: p. 111396.
32. Vulser, H., et al., *Early variations in white matter microstructure and depression outcome in adolescents with subthreshold depression.* American Journal of Psychiatry, 2018. **175**(12): p. 1255-1264.

33. Zalsman, G., et al., *Fibre tract analysis using diffusion tensor imaging reveals aberrant connectivity in a rat model of depression.* The World Journal of Biological Psychiatry, 2017. **18**(8): p. 615-623.
34. van der Marel, K., et al., *Functional and structural neural network characterization of serotonin transporter knockout rats.* PLoS One, 2013. **8**(2): p. e57780.
35. Seewoo, B.J., et al., *White matter changes following chronic restraint stress and neuromodulation: a diffusion magnetic resonance imaging study in young male rats.* Biological Psychiatry Global Open Science, 2022. **2**(2): p. 153-166.
36. Rutten-Jacobs, L.C., et al., *Genetic study of white matter integrity in UK Biobank (N= 8448) and the overlap with stroke, depression, and dementia.* Stroke, 2018.
37. Lussier, A.A., et al., *Genetic susceptibility for major depressive disorder associates with trajectories of depressive symptoms across childhood and adolescence.* Journal of Child Psychology and Psychiatry, 2021. **62**(7): p. 895-904.
38. Halldorsdottir, T., et al., *Polygenic risk: predicting depression outcomes in clinical and epidemiological cohorts of youths.* American Journal of Psychiatry, 2019. **176**(8): p. 615-625.
39. Huang, H., et al., *White matter changes in healthy adolescents at familial risk for unipolar depression: a diffusion tensor imaging study.* Neuropsychopharmacology, 2011. **36**(3): p. 684-691.
40. Sajjadian, M., et al., *Machine learning in the prediction of depression treatment outcomes: a systematic review and meta-analysis.* Psychological Medicine, 2021. **51**(16): p. 2742-2751.
41. Zhou, T., et al., *Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis.* Human brain mapping, 2019. **40**(3): p. 1001-1016.
42. Mirabnahrazam, G., et al., *Machine learning based multimodal neuroimaging genomics dementia score for predicting future conversion to alzheimer's disease.* Journal of Alzheimer's Disease, 2022. **87**(3): p. 1345-1365.
43. Guo, W., J. Wang, and S. Wang, *Deep multimodal representation learning: A survey.* Ieee Access, 2019. **7**: p. 63373-63394.
44. Koppe, G., A. Meyer-Lindenberg, and D. Durstewitz, *Deep learning for small and big data in psychiatry.* Neuropsychopharmacology, 2021. **46**(1): p. 176-190.
45. Berisha, V., et al., *Digital medicine and the curse of dimensionality.* NPJ digital medicine, 2021. **4**(1): p. 153.
46. Ricard, J., et al., *Confronting racially exclusionary practices in the acquisition and analyses of neuroimaging data.* Nature Neuroscience, 2023. **26**(1): p. 4-11.
47. Goldfarb, M.G. and D.R. Brown, *Diversifying participation: The rarity of reporting racial demographics in neuroimaging research.* NeuroImage, 2022. **254**: p. 119122.
48. Sundararajan, M., A. Taly, and Q. Yan. *Axiomatic attribution for deep networks*. in *International conference on machine learning*. 2017. PMLR.
49. Smilkov, D., et al., *Smoothgrad: removing noise by adding noise.* arXiv preprint arXiv:1706.03825, 2017.

50. Tatham, E.L., et al., *The 5-HTTLPR and BDNF polymorphisms moderate the association between uncinate fasciculus connectivity and antidepressants treatment response in major depression.* European archives of psychiatry and clinical neuroscience, 2017. **267**: p. 135-147.
51. Tatham, E.L., et al., *White matter integrity in major depressive disorder: implications of childhood trauma, 5-HTTLPR and BDNF polymorphisms.* Psychiatry Research: Neuroimaging, 2016. **253**: p. 15-25.
52. Carballedo, A., et al., *Reduced fractional anisotropy in the uncinate fasciculus in patients with major depression carrying the met-allele of the Val66Met brain-derived neurotrophic factor genotype.* American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 2012. **159**(5): p. 537-548.
53. Choi, S., et al., *Effects of a polymorphism of the neuronal amino acid transporter SLC6A15 gene on structural integrity of white matter tracts in major depressive disorder.* PloS one, 2016. **11**(10): p. e0164301.
54. Tozzi, L., et al., *Single-nucleotide polymorphism of the FKBP5 gene and childhood maltreatment as predictors of structural changes in brain areas involved in emotional processing in depression.* Neuropsychopharmacology, 2016. **41**(2): p. 487-497.
55. Hayashi, K., et al., *COMT Val158Met, but not BDNF Val66Met, is associated with white matter abnormalities of the temporal lobe in patients with first-episode, treatment-naive major depressive disorder: a diffusion tensor imaging study.* Neuropsychiatric disease and treatment, 2014: p. 1183-1190.
56. Shyn, S.I. and S.P. Hamilton, *The genetics of major depression: moving beyond the monoamine hypothesis.* Psychiatric Clinics, 2010. **33**(1): p. 125-140.
57. Holmans, P., et al., *Genomewide significant linkage to recurrent, early-onset major depressive disorder on chromosome 15q.* The American Journal of Human Genetics, 2004. **74**(6): p. 1154-1167.
58. Cole, J., et al., *White matter abnormalities and illness severity in major depressive disorder.* The British Journal of Psychiatry, 2012. **201**(1): p. 33-39.
59. Murphy, M.L. and T. Frodl, *Meta-analysis of diffusion tensor imaging studies shows altered fractional anisotropy occurring in distinct brain areas in association with depression.* Biology of mood & anxiety disorders, 2011. **1**: p. 1-12.
60. Vestergaard, M., et al., *White matter microstructure in superior longitudinal fasciculus associated with spatial working memory performance in children.* Journal of cognitive neuroscience, 2011. **23**(9): p. 2135-2146.
61. Biesbroek, J.M., et al., *Association between subcortical vascular lesion location and cognition: a voxel-based and tract-based lesion-symptom mapping study. The SMART-MR study.* PloS one, 2013. **8**(4): p. e60541.
62. Maldonado, I.L., et al., *The human cingulum: from the limbic tract to the connectionist paradigm.* Neuropsychologia, 2020. **144**: p. 107487.
63. Neave, N., et al., *The effects of discrete cingulum bundle lesions in the rat on the acquisition and performance of two tests of spatial working memory.* Behavioural brain research, 1996. **80**(1-2): p. 75-85.

64. Lichenstein, S.D., T. Verstynen, and E.E. Forbes, *Adolescent brain development and depression: a case for the importance of connectivity of the anterior cingulate cortex.* Neuroscience & Biobehavioral Reviews, 2016. **70**: p. 271-287.
65. Kaess, M., et al., *Advancing a temporal framework for understanding the biology of nonsuicidal self-injury: An expert review.* Neuroscience & Biobehavioral Reviews, 2021. **130**: p. 228-239.
66. Tian, F., et al., *The Correlation of Reduced Fractional Anisotropy in the Cingulum With Suicide Risk in Bipolar Disorder.* Front Psychiatry, 2021. **12**: p. 707622.
67. Abraham, M., et al., *Unraveling the mystery of white matter in depression: A translational perspective on recent advances.* Brain and behavior, 2022. **12**(7): p. e2629.
68. Radoeva, P.D., et al., *Systematic Review: White Matter Microstructural Organization in Adolescents With Depression.* JAACAP open, 2023.
69. Guo, W.-b., et al., *Altered white matter integrity in young adults with first-episode, treatment-naive, and treatment-responsive depression.* Neuroscience letters, 2012. **522**(2): p. 139-144.
70. Sugimoto, K., et al., *Relationship between white matter integrity and serum inflammatory cytokine levels in drug-naive patients with major depressive disorder: diffusion tensor imaging study using tract-based spatial statistics.* Translational psychiatry, 2018. **8**(1): p. 141.
71. Won, E., et al., *Association between reduced white matter integrity in the corpus callosum and serotonin transporter gene DNA methylation in medication-naive patients with major depressive disorder.* Translational Psychiatry, 2016. **6**(8): p. e866-e866.
72. Olivier, J.D., et al., *A study in male and female 5-HT transporter knockout rats: an animal model for anxiety and depression disorders.* Neuroscience, 2008. **152**(3): p. 573-84.
73. Bleys, D., et al., *Gene-environment interactions between stress and 5-HTTLPR in depression: A meta-analytic update.* Journal of affective disorders, 2018. **226**: p. 339-345.
74. Vickers, A.J. and F. Holland, *Decision curve analysis to evaluate the clinical benefit of prediction models.* The Spine Journal, 2021. **21**(10): p. 1643-1648.
75. Rice, M.E. and G.T. Harris, *Comparing effect sizes in follow-up studies: ROC Area, Cohen's d, and r.* Law and human behavior, 2005. **29**: p. 615-620.
76. Schmaal, L., et al., *Subcortical brain alterations in major depressive disorder: findings from the ENIGMA Major Depressive Disorder working group.* Molecular psychiatry, 2016. **21**(6): p. 806-812.
77. Schmaal, L., et al., *Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group.* Molecular psychiatry, 2017. **22**(6): p. 900-909.
78. Rheault, F., et al., *Common misconceptions, hidden biases and modern challenges of dMRI tractography.* J Neural Eng, 2020. **17**(1): p. 011001.
79. Calamante, F., *Track-weighted imaging methods: extracting information from a streamlines tractogram.* Magnetic Resonance Materials in Physics, Biology and Medicine, 2017. **30**(4): p. 317-335.

80. Calamante, F., et al., *A generalised framework for super-resolution track-weighted imaging.* Neuroimage, 2012. **59**(3): p. 2494-2503.
81. Casey, B.J., et al., *The adolescent brain cognitive development (ABCD) study: imaging acquisition across 21 sites.* Developmental cognitive neuroscience, 2018. **32**: p. 43-54.
82. Filipiak, P., et al., *Towards linking diffusion MRI based macro- and microstructure measures with cortico-cortical transmission in brain tumor patients.* Neuroimage, 2021. **226**: p. 117567.
83. Tournier, J.-D., et al., *MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation.* Neuroimage, 2019. **202**: p. 116137.
84. Willats, L., et al., *Quantification of track-weighted imaging (TWI): characterisation of within-subject reproducibility and between-subject variability.* Neuroimage, 2014. **87**: p. 18-31.
85. Pascual-Diaz, S., et al., *Default Mode Network structural alterations in Kocher-Monro trajectory white matter transection: A 3 and 7 tesla simulation modeling approach.* PloS one, 2019. **14**(11): p. e0224598.
86. Achenbach, T.M., *Manual for ASEBA school-age forms & profiles.* University of Vermont, Research Center for Children, Youth & Families, 2001.
87. Townsend, L., et al., *Development of three web-based computerized versions of the kiddie schedule for affective disorders and schizophrenia child psychiatric diagnostic interview: preliminary validity data.* Journal of the American Academy of Child & Adolescent Psychiatry, 2020. **59**(2): p. 309-325.
88. Romer, A.L., B. Ren, and D.A. Pizzagalli, *Brain structure relations with psychopathology trajectories in the ABCD study.* Journal of the American Academy of Child & Adolescent Psychiatry, 2023. **62**(8): p. 895-907.
89. Holt-Gosselin, B., et al., *Familial risk for depression moderates neural circuitry in healthy preadolescents to predict adolescent depression symptoms in the Adolescent Brain Cognitive Development (ABCD) Study.* Dev Cogn Neurosci, 2024. **68**: p. 101400.
90. Loughnan, R.J., et al., *Unique prediction of developmental psychopathology from genetic and familial risk.* Journal of Child Psychology and Psychiatry, 2022. **63**(12): p. 1631-1643.
91. Heo, E.-H., et al., *Validation of the center for epidemiological studies depression scale among Korean adolescents.* Psychiatry investigation, 2017. **15**(2): p. 124.
92. Radloff, L.S., *The use of the Center for Epidemiologic Studies Depression Scale in adolescents and young adults.* Journal of youth and adolescence, 1991. **20**(2): p. 149-166.
93. 이미숙, *한국판 CES-D 척도 (The Center for Epidemiologic Studies-Depression Scale) 의 요인구조분석에 대한 재검토.* 보건과 사회과학, 2002. **12**: p. 43-62.
94. Oh, K.-J. and Y.-A. Kim, *Manual for the ASEBA school-age forms.* Seoul: Huno, 2010.
95. Vilagut, G., et al., *Screening for depression in the general population with the Center for Epidemiologic Studies Depression (CES-D): a systematic review with meta-analysis.* PloS one, 2016. **11**(5): p. e0155431.

96. Henry, S.K., M.M. Grant, and K.L. Cropsey, *Determining the optimal clinical cutoff on the CES-D for depression in a community corrections sample.* Journal of affective disorders, 2018. **234**: p. 270-275.
97. Joo, Y.Y., et al., *Polygenic architecture of brain structure and function, behaviors, and psychopathologies in children.* bioRxiv, 2024: p. 2024.05.22.595444.
98. Ruan, Y., et al., *Improving polygenic prediction in ancestrally diverse populations.* Nature genetics, 2022. **54**(5): p. 573-580.
99. Meisner, A., et al., *Combined utility of 25 disease and risk factor polygenic risk scores for stratifying risk of all-cause mortality.* The American Journal of Human Genetics, 2020. **107**(3): p. 418-431.
100. Jukarainen, S., et al., *Genetic risk factors have a substantial impact on healthy life years.* Nature medicine, 2022. **28**(9): p. 1893-1901.
101. Raza, A., et al., *Enhancing brain tumor classification with transfer learning: Leveraging DenseNet121 for accurate and efficient detection.* International Journal of Imaging Systems and Technology, 2024. **34**(1): p. e22957.
102. Geng, L. *Multi-classification on the diagnosis of four early stages of Alzheimer's disease by transfer learning models*. in *Third International Conference on Artificial Intelligence and Computer Engineering (ICAICE 2022)*. 2023. SPIE.

Figures & Tables

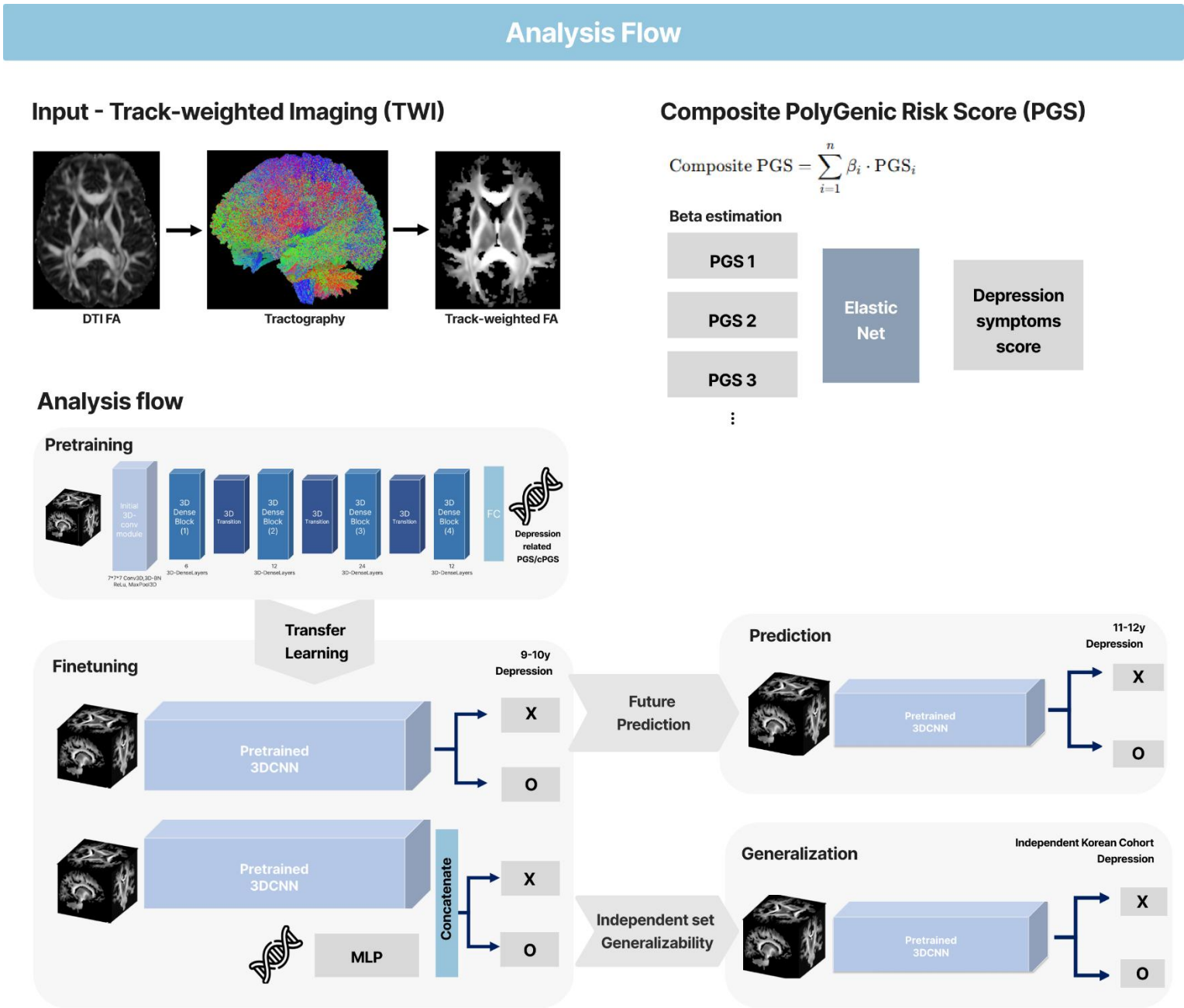


Figure 1. Analysis Flows.

This study integrates Track-Weighted FA and PGS to enhance youth depression prediction. The model is pretrained on white matter imaging and depression-related PGS using a 3D convolutional neural network. Fine-tuned model evaluates 9-10y depression classification, 2-

780 year follow-up depression prediction followed by validation in an independent Korean cohort for
781 generalizability (Multilayer perceptron layer (MLP)) .

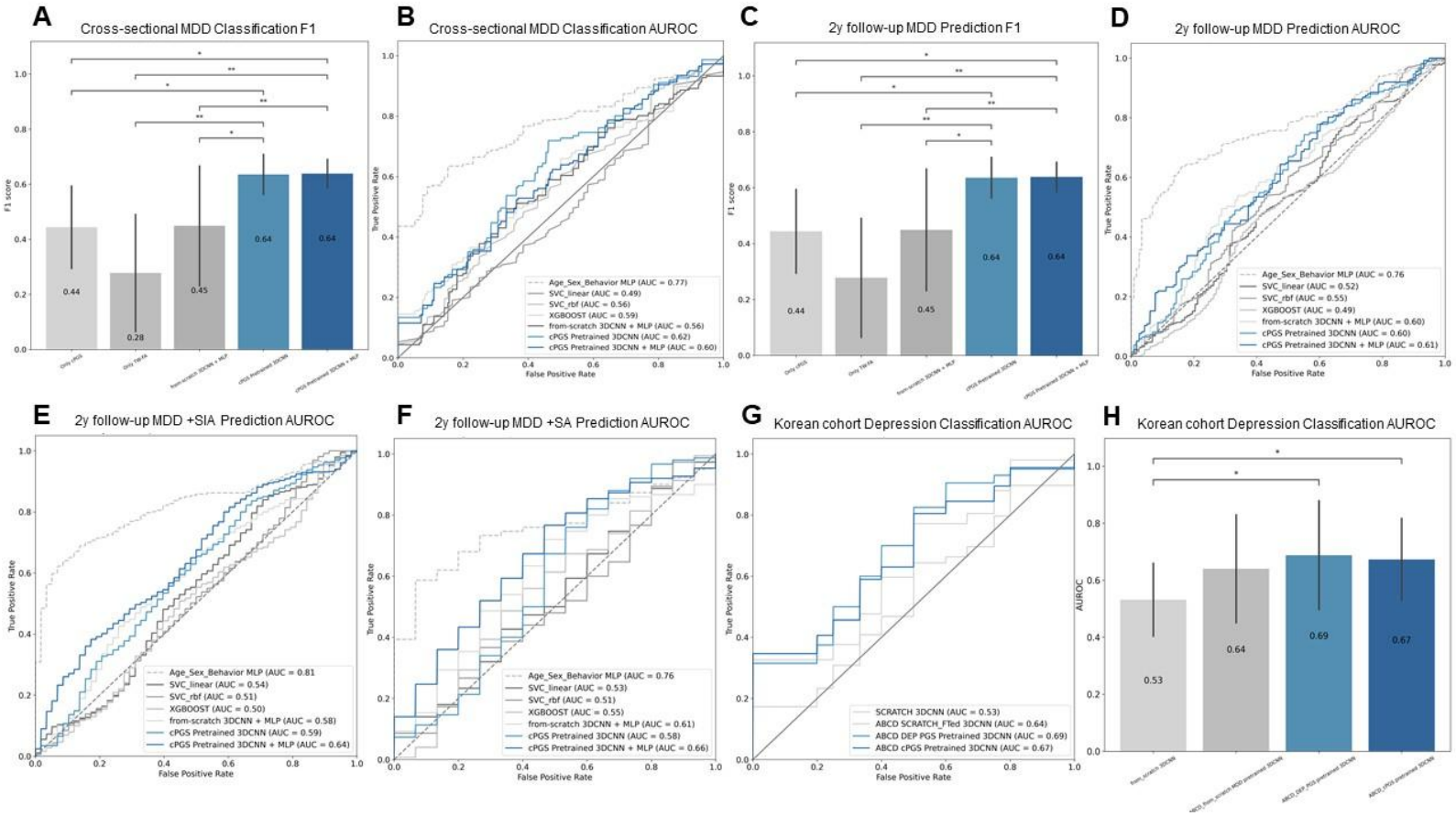


Figure 2. Depression Classification Task Performance using Composite PGS. (A) F1 score comparison for cross-sectional MDD classification, showing statistically significant performance differences across models. (B) Receiver operating characteristic (ROC) curves displaying AUROC values for cross-sectional MDD classification models. (C) F1 score comparison for 2-year follow-up MDD prediction, highlighting model performance differences. (D) ROC curves for 2-year follow-up MDD prediction. (E) ROC curves for 2-year follow-up MDD and suicidal active ideation prediction. (F) ROC curves for 2-year follow-up MDD and suicidal attempt prediction. (A-F) ABCD cohort. (G) ROC curves for depression classification in the independent Korean cohort, demonstrating external validation performance (H) AUROC comparison for depression classification, highlighting statistical significance in generalizability across models. (G-H) Korean cohort.

(*: p-value < 0.05, **: p-value <0.01, ***p-value <0.001) .

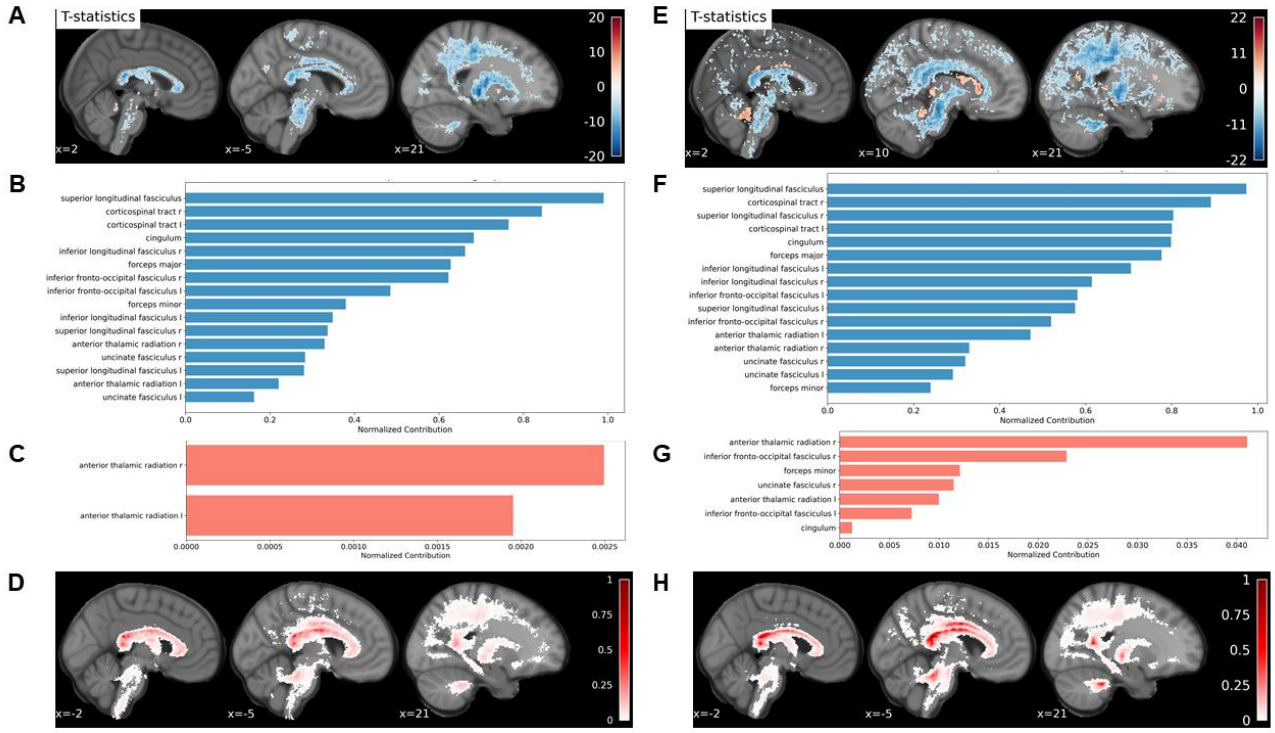


Figure3. Machine Learning Interpretations Reveal Reduced White Matter FA Associated with Youth MDD in ABCD Cohort. (A) Cross-sectional (9-10y) MDD classification explainable AI (XAI) map. Voxel-wise t-tests were performed with FDR correction ($p < 0.001$, cluster threshold = 100). Blue regions indicate voxels where lower fractional anisotropy (FA) values are associated with a higher probability of MDD. (B) Fasciculus-level contributions of voxels with significantly lower FA in the MDD group. Normalized contribution was calculated as the proportion of significant voxels (negative associations) within each fasciculus. (C) Fasciculus-level contributions of voxels where higher FA is associated with a higher probability of MDD, using the same normalization method as in (B). (D) Variability map of individual attribute score within True Positive participants. (E) XAI map for MDD prediction at 2-year follow-up in individuals with suicidal active ideation. Voxel-wise t-tests with FDR correction ($p < 0.001$, cluster threshold = 100) were applied. (F) Fasciculus-level contributions of voxels where lower FA is associated with a higher probability of MDD+SIA at 2-year follow-up, following the same analysis as (B). (G) Fasciculus-level contributions of voxels where higher FA is associated with a higher probability of MDD+SIA at 2-year follow-up, following the same analysis as (C). (H) Variability map of individual attribute score within True Positive participants.

Table 1 – Cross-sectional MDD Classification 10-fold CV Average TEST Performance Comparison with Modalities

model	input	ACC	AUROC	F1 score	stdACC	stdAUROC	std F1
ONLY_DEP PGS	DEP PGS	0.563	0.546	0.517	0.069	0.114	0.155
Only TW-FA	TW-FA	0.496	0.586	0.278	0.06	0.133	0.212
from-scratch 3DCNN	TW-FA + DEP PGS	0.533	0.54	0.484	0.084	0.146	0.223
DEP PGS pretrained 3DCNN	TW-FA	**0.585**	**0.597**	**0.629**	0.096	0.12	0.085
DEP PGS pretrained 3DCNN	TW-FA + DEP PGS	0.547	0.574	0.619	0.066	0.081	0.071
model	**input**	**ACC**	**AUROC**	**F1 score**	**stdACC**	**stdAUROC**	**std F1**
ONLY cPGS	cPGS	0.473	0.5	0.444	0.086	0.118	0.149
Only TW-FA	TW-FA	0.496	0.586	0.278	0.06	0.133	0.212
from-scratch 3DCNN	TW-FA + cPGS	0.503	0.562	0.449	0.071	0.118	0.217
cPGS pretrained 3DCNN	TW-FA	**0.577**	**0.62**	0.636	0.094	0.121	0.072
cPGS pretrained 3DCNN	TW-FA + cPGS	0.564	0.6	**0.638**	0.033	0.127	0.052

Table 2– 2-Year-Follow-up MDD Prediction 10-fold CV Average TEST Performance Comparison with Modalities

model	input	ACC	AUROC	F1 score	stdACC	stdAUROC	std F1
ONLY DEP PGS	DEP PGS	0.529	0.538	0.473	0.01	0.01	0.017
from-scratch 3DCNN	TW-FA	0.575	0.589	0.534	0.01	0.012	0.016
from-scratch 3DCNN	TW-FA + DEP PGS	**0.589**	0.607	0.565	0.01	0.014	0.02
DEP PGS pretrained 3DCNN	TW-FA	0.583	0.595	0.605	0.014	0.014	0.018
DEP PGS pretrained 3DCNN	TW-FA + DEP PGS	0.581	**0.619**	**0.623**	0.01	0.014	0.012
model	**input**	**ACC**	**AUROC**	**F1 score**	**stdACC**	**stdAUROC**	**std F1**
ONLY composite PGS	composite PGS	0.536	0.562	0.474	0.009	0.011	0.016
Only TW-FA	TW-FA	0.575	0.589	0.534	0.01	0.012	0.016
from-scratch 3DCNN	TW-FA + composite PGS	0.577	0.596	0.604	0.008	0.008	0.011
composite PGS pretrained 3DCNN	TW-FA	0.566	0.597	0.593	0.013	0.013	0.017
composite PGS pretrained 3DCNN	TW-FA + composite PGS	**0.585**	**0.61**	**0.652**	0.006	0.011	0.007

Table 3. Korean Intendent Set Depression Classification 10-fold CV Average TEST Performance

model	input	ACC	AUROC	F1	stdev ACC	stdevAUROC	stdev F1
from-scratch 3DCNN	TW-FA	0.574	0.532	0.5	0.119	0.127	0.215
ABCD_fromscratch_ Pretrained 3DCNN	TW-FA	0.615	0.64	**0.627**	0.18	0.189	0.235
ABCD_DEP PGS_ Pretrained 3DCNN	TW-FA	0.59	**0.688**	0.488	0.145	0.141	0.295
ABCD_cPGS_ Pretrained 3DCNN	TW-FA	**0.636**	0.673	0.61	0.12	0.154	0.227

Supplementary Materials

Polygenic Risk-Informed White Matter Integrity Improves Deep Learning-Based Prediction of Youth Depression

Supplementary Materials

MRI Acquisition

MRI Acquisition of ABCD Cohort

MRI data for the ABCD study was gathered from 21 research sites using Siemens Prisma, GE 750, and Philips Achieva and Ingenia 3T scanners. To maintain uniformity, scanning protocols were standardized across all sites. Detailed descriptions of the imaging protocols for both structural and dMRI have been previously published [1, 2]. The dMRI data were obtained in the axial plane with a 1.7mm isotropic resolution and a multiband acceleration factor of 3. The diffusion-weighted images included seven b=0 s/mm² frames and 96 non-collinear gradient directions, with 6 directions at b=500 s/mm², 15 directions at b=1000 s/mm², 15 directions at b=2000 s/mm², and 60 directions at b=3000 s/mm². T1-weighted images were captured using a 3D magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence at a 1mm isotropic resolution, without multiband acceleration.

The dMRI data from the ABCD study were acquired and preprocessed following a standardized protocol. For MRI quality assessment, we conducted the ABCD DAIRC's pipeline. The detailed inclusion criteria are available in the ABCD Release Notes, under 'MRI QC and Recommended Image Inclusion Criteria'. The DAIRC processed the images with several steps, including correction for eddy current distortion and head motion, adjustment of diffusion gradients, robust diffusion tensor estimation, B0 distortion correction and resampling. Comprehensive details of the dMRI preprocessing are available in the ABCD Release Notes. We employed MRtrix3 [3] for calculating the diffusion tensor metric, FA, and generating tractography. FA reflects the magnitude and direction of water diffusivity, providing insights into the organization of axons and myelin sheaths [4, 5]. FA values range from 0 to 1, indicating the degree of anisotropy of water molecules and their directional diffusion within each voxel. Probabilistic tractography was performed using constrained-spherical deconvolution (CSD) [6], with random seeding across the brain and a target streamline count of 20 million. The initial tractograms were filtered using spherical-deconvolution informed filtering (SIFT) (2:1 ratio) [7], yielding a final streamline count of 10 million, which was further down-sampled to 1 million.

The computations were conducted using supercomputers at the Argonne Leadership Computing Facility (Theta) and the Texas Advanced Computing Center (Stampede2).

MRI Acquisition of Korean Cohort

Diffusion MRI data were acquired using Siemens MAGNETOM Trio Tim 3T scanners with the following parameters: voxel size of 2.3 × 2.3 × 2.3 mm, TR of 3200 ms, TE of 110 ms, and a multi-band acceleration factor of 3. The acquisition included 66 slices in the axial plane with interleaved multi-slice mode. Diffusion-weighted images were obtained using monopolar schemes with b-values of 0 and 3000 s/mm², specifically including 20 directions at b=1000 s/mm², 30 directions at b=2000 s/mm², and 64 directions at b=3000 s/mm². T1-weighted images were collected with a 3D magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence at 1 mm isotropic resolution without multiband acceleration, providing high-resolution structural data. Probabilistic tractography was performed using CSD, with random seeding across the brain and the target streamline was 1 million. All the preprocessing steps were conducted using the Brainlife platform [8].

Participants

We utilized a large-scale neuroimaging, cognitive, and behavioral dataset from 11,868 multiethnic children as part of the Adolescent Brain and Cognitive Development (ABCD) study [1], a nationwide, longitudinal investigation into brain development and child health. For this study, we accessed release 5.1 (http://abcdstudy.org) for demographic, cognitive, psychiatric, and behavioral assessments to correct missing or erroneous values from earlier releases. We used release 2.0 for de-identified diffusion MRI neuroimaging data. We analyzed baseline data (9–10 years old) and 2-year follow-up data (11–12 years old) for psychological assessments. Participants were excluded based on several criteria: not meeting diffusion MRI quality control standards, not being fluent in English, having a history of severe traumatic brain injury, or being one of multiple births (twins or triplets) from which only one was randomly selected. Additionally, we included participants who had both genetic and diffusion MRI data. Ultimately,

a total of 5,248 participants were included in our analysis (Fig S1.A). The ABCD study was approved by institutional review boards, and informed consent and assent were obtained from all participants and their legal guardians.

We recruited a total of 180 Korean children and adolescents through collaboration with Seoul National University and Ulsan National Institute of Science and Technology. This dataset comprises multi-modal neuroimaging (T1, diffusion MRI, task-functional MRI), behavioral assessments from both participants and their caregivers and genomic data. Participants were excluded if they met any of the following criteria: (1) intellectual disabilities, (2) a history of being diagnosed with psychiatric disorders or learning disabilities, (3) lack of fluency in Korean, (4) color blindness, or (5) conditions that made normal MRI scanning difficult (e.g., presence of metallic implants or potential pregnancy). Among the recruited participants, 150 passed data quality control, which excluded individuals with issues such as excessive head motion or equipment-related problems. Of these, 108 participants who met the criteria for depression or health control groups were included in the final analysis (Fig. S1.B). Participants and their legal guardians provided written informed consent before the experiment, and all procedures were approved by the IRB of Seoul National University. (IRB No. 2112/004-006, IRB No. 2202/003-003)

Demographic variables

Demographic, Biological variables of ABCD Cohort

We used age, biological sex, and ABCD study site data at the baseline year from ABCD release 5.1. To minimize the influence of these variables, we utilized iterative stratification and propensity score matching (PSM) using these variables to match participants in both the validation and test sets across all experimental designs. We used PsmPy for PSM analysis [9].

Genetic data preprocessing

907 In our study, quality control was stringently applied to the genotyping of 733,293 SNPs
908 from saliva DNA samples using the Affymetrix NIDA Smoke Screen Array at Rutgers University
909 Cell and DNA Repository RUCDR. We excluded SNPs and samples with less than 95% call rate
910 and SNPs with a minor allele frequency below 1% using PLINK 1.90. Following this, genotypes
911 were imputed with the 1000 Genomes phase 3 v5 panel [10] and further filtered for high-quality
912 variants using strict criteria, including Hardy-Weinberg equilibrium. Additionally, we utilized
913 PC-Air [11] and PC-Relate [12] to ensure the inclusion of only genetically unrelated individuals
914 from the multi-ethnic ABCD study cohort, ultimately retaining 11,301,999 variants from 8,620
915 participants. For further detail on these processes, see Joo et al [13].

916

917 **Model Optimization and Evaluation**

918 *Pretraining with PGS*

919 Participants of pretraining set were partitioned into train, validation, test datasets with 10-
920 fold iterative stratification to ensure there was no significant difference in distribution of
921 biological sex, ABCD study site and age (month) and the prediction target outcome, which was
922 PGS. Train, validation, and test were divided in a ratio of 7:2:1. Hyperparameter tuning was
923 performed within the following ranges: learning rate (lr) = {0.01, 0.001}, weight decay (wd) =
924 {0.01, 0.001,0.001,0.00001}, dropout = {0.1, 0.2} and compared validation performance metric
925 to select the best model.

926 *Cross-sectional MDD Prediction*

927 Participants were partitioned with 10-fold iterative stratification to ensure there was no
928 significant difference in distribution of biological sex, ABCD study site and age (month) and the
929 prediction target outcome. Train, validation, and test were divided in a ratio of 7:2:1. We
930 conducted PSM to match the participants in the folds using same covariate variables.
931 Hyperparameter tuning was performed within the following ranges: learning rate (lr) = {0.01,
932 0.001, 0.0001, 0.00001}, weight decay (wd) = {0.01, 0.001}, dropout = {0.1, 0.2} and compared
933 validation performance metric to select the best model. All the metrics are reported as average
934 values across all repeated settings. The pretrain and finetune sets, while held out from one

another, share similar data distributions, and the outcomes for pretraining and finetuning are closely aligned. Also, considering the relatively small size of the finetune set, we chose to freeze the pretrained model's parameters, allowing only the final layer to be updated during finetuning for fully utilize pre-trained knowledge [14].

2-year follow-up MDD / MDD & Suicidal Behavior Prediction

The dataset for 2-year follow-up predictions excluded participants from cross-sectional tasks. Predictions for future depression KSADS diagnosis were made using an ensemble of the 10 models generated in the cross-sectional MDD KSADS diagnosis classification task. For statistical performance comparison, as previously described, 10 randomly matched datasets were inferred, and the reported performance metrics represent the average across these datasets. The 2-year MDD KSADS diagnosis prediction involved 10 datasets comprising a total of 236 participants. For suicidal behavior prediction, the same strategy was applied, resulting in 10 datasets including 116 participants with MDD and suicidal active ideation at the 2-year follow-up, and 30 participants with MDD and a history of suicide attempts. However, we ensured that none of the participants overlapped with pretraining nor finetune sets.

Transfer Learning to Independent Dataset

Participants of Korean datasets were also partitioned through 10-fold iterative stratification to ensure there was no significant difference in distribution of biological sex, age (month) and the prediction target outcome. Train, validation, and test were divided in a ratio of 7:2:1. We conducted PSM to match the participants in the folds using same covariate variables. Hyperparameter tuning was performed within the following ranges: learning rate (lr) = {0.01, 0.001, 0.0001, 0.00001}, weight decay (wd) = {0.01, 0.001}, dropout = {0.1, 0.2} and compared validation performance metric to select the best model. All the metrics are reported in average value of all repeated settings. Considering the ethnic difference between the two datasets, we chose to unfreeze the ABCD MDD classifier model's parameters.

963 **Control Analyses**

964 Additional control analyses confirmed the robustness of our approach. First, PGS-specific
965 pretraining demonstrated superior classification accuracy compared to pretraining with
966 behavioral, family history of depression, and cognitive variables (Table S9). Both PGS-
967 pretrained models significantly outperformed the family history-pretrained model ($p < 0.05$) and
968 the total intelligence-pretrained model in accuracy ($p < 0.02$). The cPGS-pretrained model also
969 showed a marginally significant improvement over the CBCL total behavioral problems-
970 pretrained model in F1 score ($p < 0.06$, Table S28, Fig. S4A–D).

971 Second, we examined the role of microstructural versus macrostructural information in
972 MDD classification. Models using macrostructural-only inputs (e.g., white matter shape and
973 volume) performed significantly worse in both cross-sectional and 2-year follow-up MDD
974 prediction compared to TW-FA-based models, with statistically significant F1 score differences
975 ($p < 0.05$, Table S9, Fig. S4E, F).

976 Third, we conducted comparisons between MDD subgroups, examining differences
977 between those with and without suicidal behaviors and healthy controls (Table S10, S11, Fig.
978 S5A, B). Lastly, we performed additional comparisons incorporating behavioral, age, and sex
979 variables alongside neuroimaging and genetic data, evaluating their combined predictive
980 performance relative to symptom-based models alone (Fig. S6A–D).

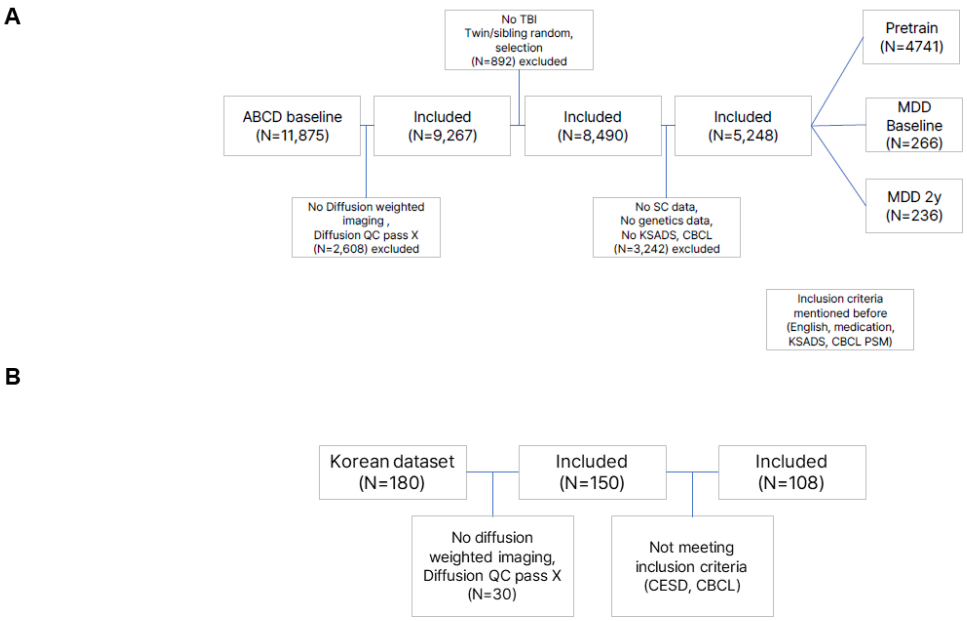
981

982

983

Supplementary Figures

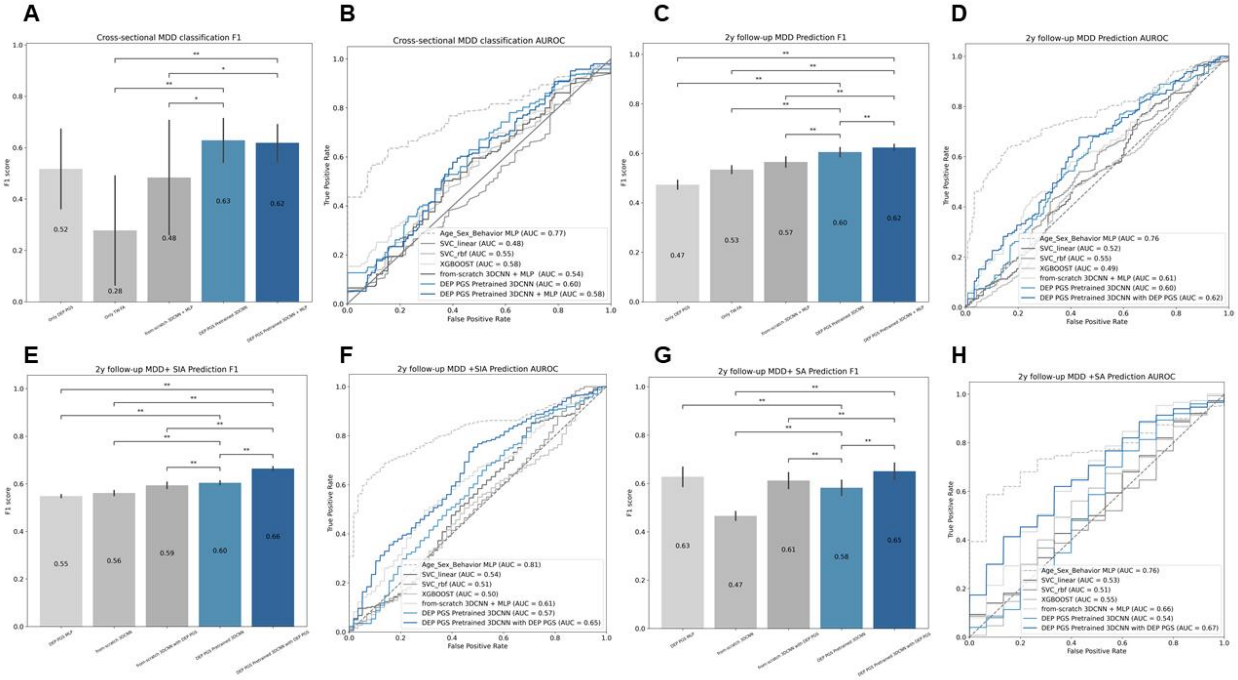
Figure-S1 – Participants inclusion / exclusion



Supplementary figure 1. Participants inclusion / exclusion process

(A) ABCD participants inclusion / exclusion process, (B) Korean dataset participants inclusion / exclusion process

Figure-S2 – ABCD Cross-sectional MDD classification & 2-Year- Follow- Up MDD prediction -DEP PGS included models Performance

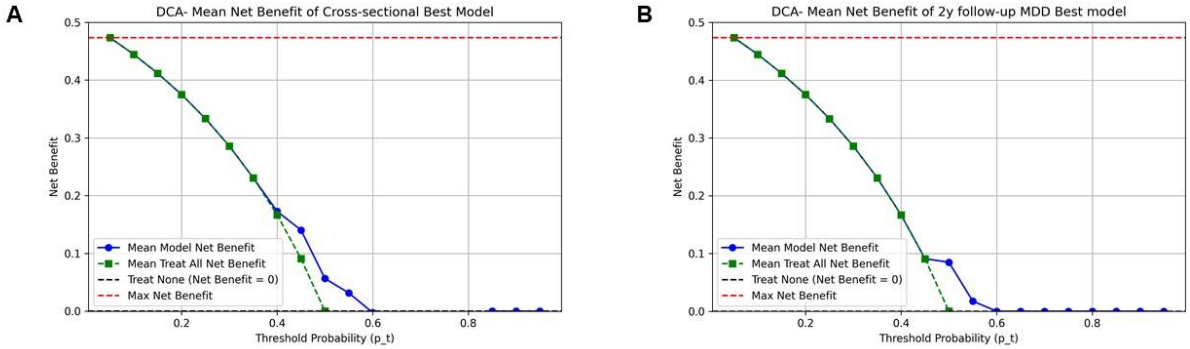


Supplementary figure 2. MDD classification task performance using Depression PGS

(A) F1 score comparison for cross-sectional MDD classification, showing statistically significant performance differences across models. (B) Receiver operating characteristic (ROC) curves displaying AUROC values for cross-sectional MDD classification models. (C) F1 score comparison for 2-year follow-up MDD prediction, highlighting model performance differences. (D) ROC curves for 2-year follow-up MDD prediction. (E) F1 score comparison for 2-year follow-up MDD and suicidal active ideation prediction. (F) ROC curves for 2-year follow-up MDD and suicidal active ideation prediction. (G) F1 score comparison for 2-year follow-up MDD and suicidal attempt prediction (H) ROC curves for 2-year follow-up MDD and suicidal attempt prediction.

(*: p-value < 0.05, **: p-value <0.01, ***p-value <0.0001)

1002 Figure-S3– Decision curve analysis

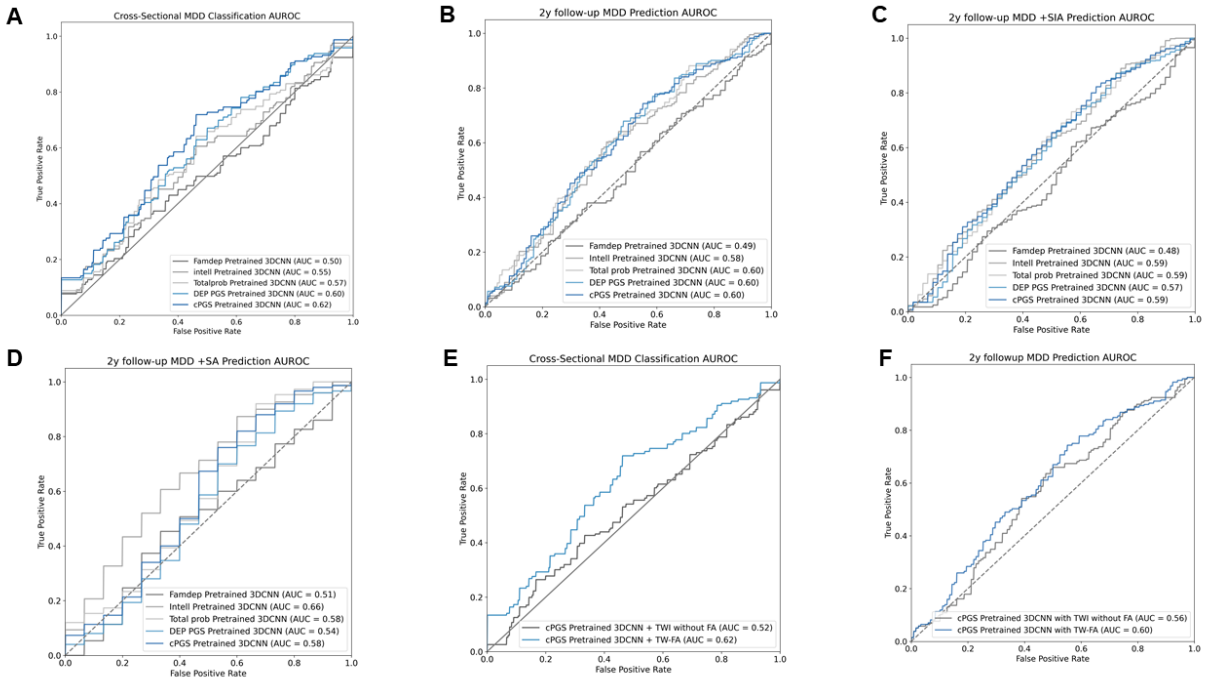


1003

1004 Supplementary figure 3. Decision curve analysis
1005 (A)Decision curve analysis for cross-sectional MDD Classification.
1006 (B) Decision curve analysis for cross-sectional 2y follow-up MDD Prediction.

1007

Figure-S4– Control analysis - Pretraining with alternative variables

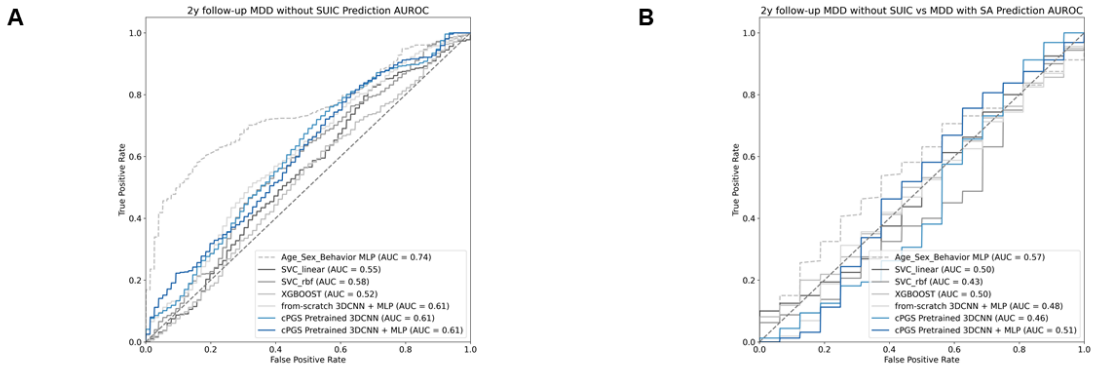


Supplementary figure 4. Pretraining with alternative variables

(A–D) AUROC curves comparing pretraining with family history of depression (FamDep), intelligence (Intell), and total problem scores (CBCL) instead of PGS for MDD classification and prediction.

(E–F) AUROC results for models trained using track-weighted imaging (TWI) features alone, excluding FA values.

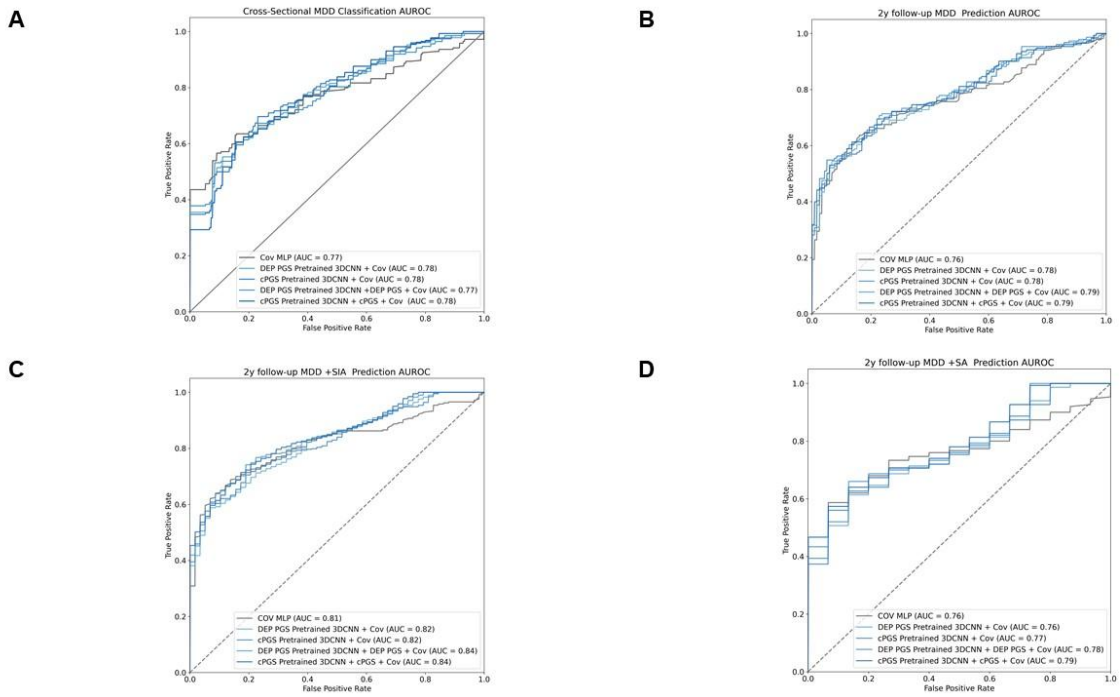
1016 Figure-S5– Control analysis – Classification performance for MDD with and without suicidal behavior



1017 Supplementary figure 5. Classification performance for MDD with and without suicidal behavior
1018 (A) AUROC curves for distinguishing health controls from MDD without suicidal behavior.
1019 (B) AUROC curves for differentiating MDD with suicidal attempts from MDD without suicidal behavior.

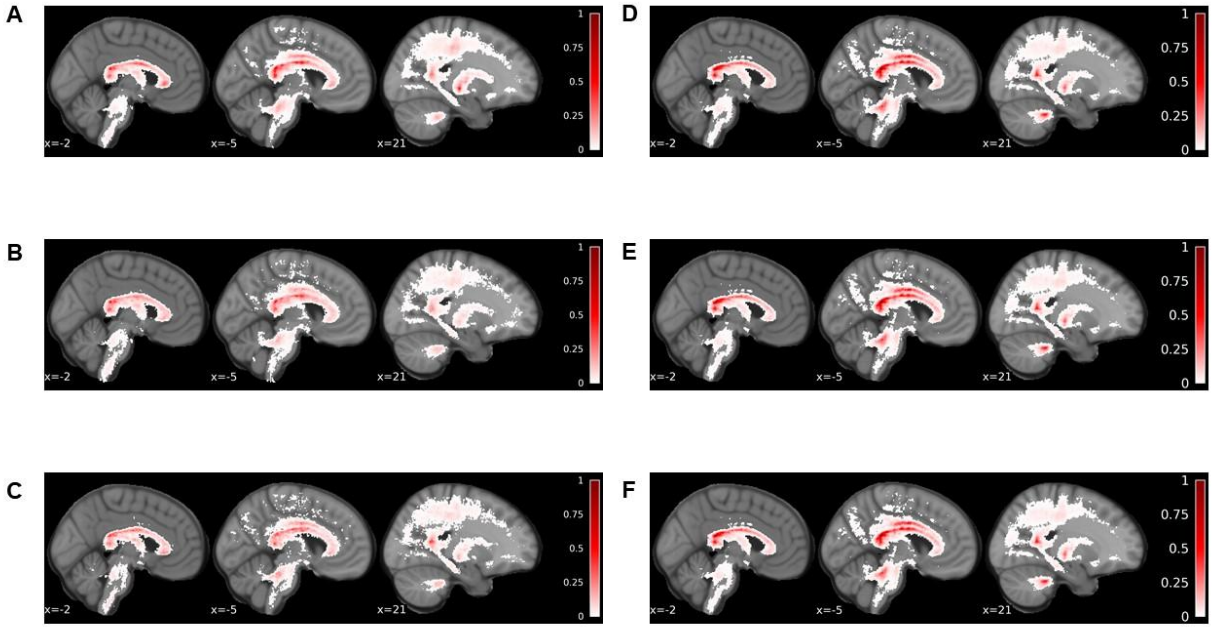
1020

Figure-S6– Control analysis – Classification performance for MDD including age, sex, behavioral variables



Supplementary figure 6. Classification performance for MDD including age, sex, behavioral variables (A) ROC curves displaying AUROC values for cross-sectional MDD classification models. (B) ROC curves for 2-year follow-up MDD prediction. (C) ROC curves for 2-year follow-up MDD and suicidal active ideation prediction. (D) ROC curves for 2-year follow-up MDD and suicidal attempt prediction.

Figure-S7–XAI variability map and individual XAI map



Supplementary figure 7. XAI variability

(A-C) Cross-sectional MDD classification (D-F) 2-y Follow-up MDD + suicidal ideation Prediction

(A) Voxels showing the top 5% variability in True Positive and True Negative individuals. (B) Voxels showing the top 5% variability in True Positive individuals. (C) Voxels showing the top 5% variability in True Negative individuals. (D) Voxels showing the top 5% variability in True Positive and True Negative individuals. (E) Voxels showing the top 5% variability in True Positive individuals. (F) Voxels showing the top 5% variability in True Negative individuals.

Supplementary Tables

Supplementary Table 1– ABCD Demographics of each Datasets

ABCD		**MDD (KSADS)**			
	Covariate	**0**	**1**	**p-value**	**statistics**
Test	sex	M=251/W=250	M=61/W=72	0.4408	0.594299
Chi_square	abcd site (max N)	16	4	0.6514	17.961729
Chi_square	race_ethnicity			0.1269	7.174994
Chi_square	white (1)	278	57		
	black (2)	58	22		
	hispanic (3)	115	39		
	asian (4)	6	2		
	Other	44	13		
T-test	age (mean)	119.311377	119.744361	0.5464	0.604158
ABCD		**MDD_2y(KSADS)**			
	Covariate	**0**	**1**	**p-value**	**statistics**
Test	sex	M=59/W=59	M=62/W=92	0.1392	2.187298
Chi_square	abcd site (max N)	16	16	0.2823	23.132551
Chi_square	race_ethnicity			0.5494	3.050694
Chi_square	white (1)	66	76		
	black (2)	13	24		
	hispanic (3)	27	38		
	asian (4)	1	0		
	Other	11	16		
T-test	age (mean)	144.466102	145.272727	0.4105	0.824294
ABCD		**MDD_2y_SIA(KSADS)**			
	Covariate	**0**	**1**	**p-value**	**statistics**
Test	sex	M=37/W=44	M=20/W=38	0.2507	1.319151
Chi_square	abcd site (max N)	16	16	0.4832	19.599870
Chi_square	race_ethnicity			0.5690	2.933352
Chi_square	white (1)	43	27		
	black (2)	9	10		
	hispanic (3)	21	13		

	asian (4)	1	0		
	Other	7	8		
T-test	age (mean)	143.814815	146.948276	0.0280	2.224974
ABCD		**MDD_2y_SA(KSADS)**			
	Covariate	**0**	**1**	**p-value**	**statistics**
Test	sex	M=32/W=39	M=4/W=11	0.3055	1.050169
Chi_square	abcd site (max N)	20	20	0.5991	17.822658
Chi_square	race_ethnicity			0.1166	7.390975
Chi_square	white (1)	37	8		
	black (2)	7	5		
	hispanic (3)	19	1		
	asian (4)	1	0		
	Other	7	1		
T-test	age (mean)	144.507042	145.333333	0.7214	0.361603

Supplementary Table 2– Korean Dataset Demographics

Korean dataset		**CES-D**			
	Covariate	**0**	**1**	**p-value**	**statistics**
Chi_square	sex	M=33/W=34	M=17/W=24	0.5558	0.347062
t-test	age (mean)	13.164179	14.219512	0.0055	2.841068

Supplementary Table 3– Cross-sectional MDD classification 10-fold CV Average TEST Performance Comparison with Machine Learning Models

model	**input**	**ACC**	**AUROC**	**F1 score**	**stdACC**	**stdAUROC**	**std F1**
SCV_linear	SC + DEP PGS	0.533	0.482	0.515	0.065	0.067	0.094
SCV_rbf	SC + DEP PGS	0.578	0.549	0.578	0.095	0.121	0.126
XGBOOST	SC + DEP PGS	0.555	0.576	0.528	0.07	0.102	0.088
from-scratch 3DCNN	TW-FA + DEP PGS	0.533	0.54	0.484	0.084	0.146	0.223
DEP PGS pretrained 3DCNN	TW-FA	**0.585**	**0.597**	**0.629**	0.096	0.12	0.085
DEP PGS pretrained 3DCNN	TW-FA + DEP PGS	0.547	0.58	0.619	0.066	0.081	0.071

model	input	ACC	AUROC	F1 score	stdACC	stdAUROC	std F1
SCV_linear	SC + cPGS	0.514	0.486	0.506	0.072	0.061	0.094
SCV_rbf	SC + cPGS	0.563	0.557	0.568	0.105	0.124	0.131
XGBOOST	SC + cPGS	0.566	0.588	0.536	0.073	0.092	0.093
from-scratch 3DCNN	TW-FA + cPGS	0.503	0.562	0.449	0.071	0.118	0.217
cPGS pretrained 3DCNN	TW-FA	**0.577**	**0.62**	0.636	0.094	0.121	0.072
cPGS pretrained 3DCNN	TW-FA + cPGS	0.564	0.6	**0.638**	0.033	0.127	0.052

Supplementary Table 4– 2-Year-Followup MDD Prediction 10-fold CV Average TEST Performance Comparison with Machine Learning Models

model	input	ACC	AUROC	F1 score	stdACC	stdAUROC	std F1
SCV linear	SC + DEP PGS	0.498	0.524	0.121	0.007	0.01	0.015
SCV rbf	SC + DEP PGS	0.497	0.548	0.348	0.009	0.01	0.019
XGBOOST	SC + DEP PGS	0.497	0.49	0.469	0.02	0.011	0.018
from-scratch 3D-CNN	TW-FA + DEP PGS	**0.589**	0.607	0.565	0.01	0.014	0.02
DEP PGS pretrained 3D-CNN	TW-FA	0.583	0.595	0.605	0.014	0.014	0.018
DEP PGS pretrained 3D-CNN	TW-FA + DEP PGS	0.581	**0.619**	**0.623**	0.01	0.014	0.012
model	**input**	**ACC**	**AUROC**	**F1 score**	**stdACC**	**stdAUROC**	**std F1**
SCV linear	SC + cPGS	0.493	0.522	0.106	0.009	0.01	0.018
SCV rbf	SC + cPGS	0.498	0.548	0.348	0.007	0.01	0.017
XGBOOST	SC + cPGS	0.495	0.494	0.472	0.014	0.009	0.016
from-scratch 3D-CNN	TW-FA + cPGS	0.577	0.596	0.604	0.008	0.008	0.011
cPGS pretrained 3D-CNN	TW-FA	0.566	0.597	0.593	0.013	0.013	0.017
cPGS pretrained 3D-CNN	TW-FA + cPGS	**0.585**	**0.61**	**0.652**	0.006	0.011	0.007

Supplementary Table 5– 2-Year-Followup MDD + Suicidal Active Ideation Prediction 10-fold CV Average TEST Performance Comparison with Modalities

model	input	ACC	AUROC	F1 score	stdACC	stdAUROC	std F1
ONLY DEP PGS	DEP PGS	0.588	0.614	0.548	0.01	0.016	0.006
from-scratch 3D-CNN	TW-FA	0.582	0.562	0.561	0.017	0.021	0.01
from-scratch 3D-CNN	TW-FA + DEP PGS	0.598	0.609	0.594	0.022	0.022	0.013
DEP PGS pretrained 3D-CNN	TW-FA	0.571	0.574	0.604	0.015	0.024	0.008
DEP PGS pretrained 3D-CNN	TW-FA + DEP PGS	**0.607**	**0.651**	**0.664**	0.015	0.021	0.008
model	**input**	**ACC**	**AUROC**	**F1 score**	**stdACC**	**stdAUROC**	**std F1**
ONLY cPGS	cPGS	0.577	0.628	0.514	0.014	0.015	0.008
Only TW-FA	TW-FA	0.582	0.562	0.561	0.017	0.021	0.01
from-scratch 3D-CNN	TW-FA + cPGS	0.586	0.581	0.631	0.02	0.017	0.011
cPGS pretrained 3D-CNN	TW-FA	0.564	0.591	0.607	0.018	0.025	0.01
cPGS pretrained 3D-CNN	TW-FA + cPGS	**0.601**	**0.64**	**0.675**	0.013	0.02	0.007

Supplementary Table 6– 2-Year-Followup MDD + Suicidal Active Ideation Prediction 10-fold CV Average TEST Performance Comparison with Machine Learning Models

model	input	ACC	AUROC	F1 score	stdACC	stdAUROC	std F1
SCV linear	SC + DEP PGS	0.521	0.535	0.17	0.011	0.024	0.018
SCV rbf	SC + DEP PGS	0.487	0.51	0.352	0.01	0.017	0.006
XGBOOST	SC + DEP PGS	0.511	0.502	0.501	0.02	0.013	0.022
from-scratch 3D-CNN	TW-FA + DEP PGS	0.598	0.609	0.594	0.022	0.022	0.013
DEP PGS pretrained 3D-CNN	TW-FA	0.571	0.574	0.604	0.015	0.024	0.008
DEP PGS pretrained 3D-CNN	TW-FA + DEP PGS	**0.607**	**0.651**	**0.664**	0.015	0.021	0.008
model	**input**	**ACC**	**AUROC**	**F1 score**	**stdACC**	**stdAUROC**	**std F1**
SCV linear	SC + cPGS	0.52	0.536	0.167	0.011	0.025	0.014
SCV rbf	SC + cPGS	0.486	0.507	0.352	0.014	0.018	0.009
XGBOOST	SC + cPGS	0.507	0.496	0.502	0.021	0.01	0.021
from-scratch 3D-CNN	TW-FA + cPGS	0.586	0.581	0.631	0.02	0.017	0.011

cPGS pretrained 3D-CNN	TW-FA	0.564	0.591	0.607	0.018	0.025	0.01
cPGS pretrained 3D-CNN	TW-FA + cPGS	**0.601**	**0.64**	**0.675**	0.013	0.02	0.007

Supplementary Table 7– 2-Year-Followup MDD + Suicidal attempt Prediction 10-fold CV Average TEST Performance Comparison with Modalities

model	**input**	**ACC**	**AUROC**	**F1 score**	**stdACC**	**stdAUROC**	**std F1**
ONLY DEP PGS	DEP PGS	**0.64**	**0.716**	0.628	0.063	0.057	0.04
from-scratch 3D-CNN	TW-FA	0.54	0.573	0.466	0.033	0.062	0.018
from-scratch 3D-CNN	TW-FA + DEP PGS	0.617	0.661	0.612	0.052	0.074	0.033
DEP PGS pretrained 3D-CNN	TW-FA	0.567	0.543	0.582	0.054	0.081	0.031
DEP PGS pretrained 3D-CNN	TW-FA + DEP PGS	0.603	0.67	**0.651**	0.055	0.076	0.034
model	**input**	**ACC**	**AUROC**	**F1 score**	**stdACC**	**stdAUROC**	**std F1**
ONLY cPGS	cPGS	**0.633**	0.664	0.622	0.049	0.051	0.031
Only TW-FA	TW-FA	0.54	0.573	0.466	0.033	0.062	0.018
from-scratch 3D-CNN	TW-FA + cPGS	0.607	0.611	0.646	0.049	0.061	0.026
cPGS pretrained 3D-CNN	TW-FA	0.62	0.585	0.661	0.064	0.097	0.038
cPGS pretrained 3D-CNN	TW-FA + cPGS	0.63	**0.665**	**0.702**	0.06	0.071	0.034

Supplementary Table 8- 2-Year-Followup MDD + Suicidal attempt Prediction 10-fold CV Average TEST Performance Comparison with Machine Learning Models

model	**input**	**ACC**	**AUROC**	**F1 score**	**stdACC**	**stdAUROC**	**std F1**
SCV linear	SC + DEP PGS	0.499	0.528	0.147	0.029	0.063	0.049
SCV rbf	SC + DEP PGS	0.54	0.511	0.419	0.042	0.059	0.057
XGBOOST	SC + DEP PGS	0.541	0.545	0.556	0.043	0.063	0.07
from-scratch 3D-CNN	TW-FA + DEP PGS	**0.617**	0.661	0.612	0.052	0.074	0.033
DEP PGS pretrained 3D-CNN	TW-FA	0.567	0.543	0.582	0.054	0.081	0.031

DEP PGS pretrained 3D-CNN	TW-FA + DEP PGS	0.603	**0.67**	**0.651**	0.055	0.076	0.034
model	**input**	**ACC**	**AUROC**	**F1 score**	**stdACC**	**stdAUROC**	**std F1**
SCV linear	SC + cPGS	0.499	0.531	0.147	0.029	0.063	0.047
SCV rbf	SC + cPGS	0.54	0.512	0.443	0.027	0.058	0.054
XGBOOST	SC + cPGS	0.546	0.55	0.547	0.058	0.055	0.061
from-scratch 3D-CNN	TW-FA + cPGS	0.607	0.611	0.646	0.049	0.061	0.026
cPGS pretrained 3D-CNN	TW-FA	0.62	0.585	0.661	0.064	0.097	0.038
cPGS pretrained 3D-CNN	TW-FA + cPGS	**0.63**	**0.665**	**0.702**	0.06	0.071	0.034

Supplementary Table 9– Control Analysis Performance

model	**input**	**ACC**	**AUROC**	**F1 score**	**stdACC**	**stdAUROC**	**std F1**
cPGS pretrained 3DCNN	TW-intensity image (tractogram)	0.513	0.522	0.354	0.059	0.095	0.262
CBCL total behavioral problems pretrained 3DCNN	TW-FA	0.555	0.572	0.488	0.096	0.137	0.187
Total intelligence pretrained 3DCNN	TW-FA	0.492	0.553	0.583	0.06	0.102	0.083
Family Depression history pretrained 3DCNN	TW-FA	0.514	0.498	0.460	0.056	0.145	0.196

Supplementary Table 10– 2-Year-Followup MDD without Suicidal behavior Prediction 10-fold CV Average TEST Performance Comparison with Modalities

model	**input**	**ACC**	**AUROC**	**F1 score**	**stdACC**	**stdAUROC**	**std F1**
ONLY_DEP PGS	DEP PGS	0.491	0.499	0.402	0.013	0.02	0.006

from-scratch 3DCNN	TW-FA	0.578	0.611	0.536	0.017	0.025	0.01
from-scratch 3DCNN	TW-FA + DEP PGS	0.576	**0.615**	0.541	0.015	0.025	0.009
DEP PGS pretrained 3DCNN	TW-FA	**0.587**	0.612	**0.61**	0.019	0.022	0.011
DEP PGS pretrained 3DCNN	TW-FA + DEP PGS	0.557	0.608	0.593	**0.02**	**0.022**	**0.011**
model	**input**	**ACC**	**AUROC**	**F1 score**	**stdACC**	**stdAUROC**	**std F1**
ONLY cPGS	cPGS	0.506	0.532	0.427	0.015	0.021	0.008
Only TW-FA	TW-FA	0.578	0.611	0.536	0.017	0.025	0.01
from-scratch 3DCNN	TW-FA + cPGS	0.577	0.607	0.599	0.02	0.024	0.011
cPGS pretrained 3DCNN	TW-FA	0.575	0.611	0.598	0.024	0.025	0.013
cPGS pretrained 3DCNN	TW-FA + cPGS	**0.593**	**0.612**	**0.66**	**0.022**	**0.024**	**0.012**

Supplementary Table 11- 2-Year-Followup MDD without Suicidal behavior Prediction 10-fold CV Average TEST Performance Comparison with Machine Learning Models

model	**input**	**ACC**	**AUROC**	**F1 score**	**stdACC**	**stdAUROC**	**std F1**
SCV_linear	SC + DEP PGS	0.478	0.547	0.057	0.013	0.017	0.019
SCV_rbf	SC + DEP PGS	0.505	0.581	0.34	0.016	0.017	0.026
XGBOOST	SC + DEP PGS	0.526	0.521	0.5	0.019	0.019	0.027
from-scratch 3DCNN	TW-FA + DEP PGS	0.576	**0.615**	0.541	0.015	0.025	0.009
DEP PGS pretrained 3DCNN	TW-FA	**0.587**	0.612	**0.61**	0.019	0.022	0.011
DEP PGS pretrained 3DCNN	TW-FA + DEP PGS	0.557	0.608	0.593	**0.02**	**0.022**	**0.011**
model	**input**	**ACC**	**AUROC**	**F1 score**	**stdACC**	**stdAUROC**	**std F1**

SCV_linear	SC + cPGS	0.483	0.545	0.064	0.012	0.018	0.024
SCV_rbf	SC + cPGS	0.506	0.582	0.341	0.017	0.017	0.027
XGBOOST	SC + cPGS	0.525	0.523	0.497	0.026	0.022	0.037
from-scratch 3DCNN	TW-FA + cPGS	0.577	0.607	0.599	0.02	0.024	0.011
cPGS pretrained 3DCNN	TW-FA	0.575	**0.609**	0.598	0.024	0.025	0.013
cPGS pretrained 3DCNN	TW-FA + cPGS	**0.593**	0.607	**0.66**	**0.022**	**0.024**	**0.012**

All the Wilcoxon rank sum test analysis were conducted through python, utilizing *statsmodels*.

Supplementary Table 12 – Cross-sectional MDD Classification Model Comparison Modalities (DEP PGS)

ACC				AUROC				F1			
Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
DEP PGS MLP	DEP PGS Pretrained 3DCNN	12	0.213524	DEP PGS MLP	DEP PGS Pretrained 3DCNN	20	0.492188	DEP PGS MLP	DEP PGS Pretrained 3DCNN	11	0.105469
DEP PGS MLP	DEP PGS Pretrained 3DCNN with DEP PGS	17	0.51467	DEP PGS MLP	DEP PGS Pretrained 3DCNN with DEP PGS	26	0.921875	DEP PGS MLP	DEP PGS Pretrained 3DCNN with DEP PGS	16	0.275391
from-scratch 3DCNN	DEP PGS Pretrained 3DCNN	4	0.028402	from-scratch 3DCNN	DEP PGS Pretrained 3DCNN	19	0.678402	from-scratch 3DCNN	DEP PGS Pretrained 3DCNN	0	0.001953
from-scratch 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	4	0.090392	from-scratch 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	27	1	from-scratch 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.007686
from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	1	0.017154	from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	10	0.083984	from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	5	0.019531
from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	8	0.600179	from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	17	0.322266	from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	5	0.019531
DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	8	0.310494	DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	24	0.769531	DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	16	0.441268

Supplementary Table 13 – Cross-sectional MDD Classification Model Comparison Modalities (cPGS)

ACC				AUROC				F1			

Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
cPGS MLP	cPGS Pretrained 3DCNN	10	0.138641	cPGS MLP	cPGS Pretrained 3DCNN	10	0.083984	cPGS MLP	cPGS Pretrained 3DCNN	6	0.027344
cPGS MLP	cPGS Pretrained 3DCNN with cPGS	9	0.064453	cPGS MLP	cPGS Pretrained 3DCNN with cPGS	12	0.130859	cPGS MLP	cPGS Pretrained 3DCNN with cPGS	6	0.027344
from-scratch 3DCNN	cPGS Pretrained 3DCNN	11	0.105469	from-scratch 3DCNN	cPGS Pretrained 3DCNN	16	0.275391	from-scratch 3DCNN	cPGS Pretrained 3DCNN	0	0.001953
from-scratch 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.011616	from-scratch 3DCNN	cPGS Pretrained 3DCNN with cPGS	17	0.51467	from-scratch 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953
from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	5	0.068704	from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	15	0.232422	from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	4	0.028402
from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	0	0.011616	from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	11	0.173071	from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	2	0.005859
cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	21.5	0.905613	cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	25	0.845703	cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	13	0.260393

Supplementary Table 14 – Cross-sectional MDD Classification Model Comparison_Machine Learning (DEP PGS)

ACC				AUROC				F1			
Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
SVC_linear	DEP PGS Pretrained 3DCNN	10	0.138641	SVC_linear	DEP PGS Pretrained 3DCNN	11	0.1054687	SVC_linear	DEP PGS Pretrained 3DCNN	5	0.019531
SVC_linear	DEP PGS Pretrained 3DCNN with DEP PGS	20	0.767097	SVC_linear	DEP PGS Pretrained 3DCNN with DEP PGS	2	0.015156	SVC_linear	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
SVC_rbf	DEP PGS Pretrained 3DCNN	24	0.769531	SVC_rbf	DEP PGS Pretrained 3DCNN	16	0.275391	SVC_rbf	DEP PGS Pretrained 3DCNN	15	0.232422
SVC_rbf	DEP PGS Pretrained 3DCNN with DEP PGS	18	0.375	SVC_rbf	DEP PGS Pretrained 3DCNN with DEP PGS	24.5	0.769531	SVC_rbf	DEP PGS Pretrained 3DCNN with DEP PGS	18	0.375

XGBOOST	DEP PGS Pretrained 3DCNN	13	0.259974	XGBOOST	DEP PGS Pretrained 3DCNN	16	0.2753906	XGBOOST	DEP PGS Pretrained 3DCNN	6	0.027344
XGBOOST	DEP PGS Pretrained 3DCNN with DEP PGS	17	0.888638	XGBOOST	DEP PGS Pretrained 3DCNN with DEP PGS	19	0.678402	XGBOOST	DEP PGS Pretrained 3DCNN with DEP PGS	7	0.037109
scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	1	0.017154	scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	14	0.193359	scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	5	0.019531
scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	8	0.600179	scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	16	0.275391	scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	5	0.019531
DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	8	0.310494	DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	22	0.952765	DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	16	0.441268

Supplementary Table 15 – Cross-sectional MDD Classification Model Comparison_Machine Learning (cPGS)

ACC				AUROC				F1			
Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
SVC_linear	cPGS Pretrained 3DCNN	15	0.232422	SVC_linear	cPGS Pretrained 3DCNN	6	0.027344	SVC_linear	cPGS Pretrained 3DCNN	2	0.015156
SVC_linear	cPGS Pretrained 3DCNN with cPGS	24	0.769531	SVC_linear	cPGS Pretrained 3DCNN with cPGS	7	0.066316	SVC_linear	cPGS Pretrained 3DCNN with cPGS	1	0.003906
SVC_rbf	cPGS Pretrained 3DCNN	27	1	SVC_rbf	cPGS Pretrained 3DCNN	16	0.232422	SVC_rbf	cPGS Pretrained 3DCNN	15	0.232422
SVC_rbf	cPGS Pretrained 3DCNN with cPGS	12	0.130859	SVC_rbf	cPGS Pretrained 3DCNN with cPGS	23	0.695313	SVC_rbf	cPGS Pretrained 3DCNN with cPGS	9	0.064453
XGBOOST	cPGS Pretrained 3DCNN	26	0.921875	XGBOOST	cPGS Pretrained 3DCNN	16	0.275391	XGBOOST	cPGS Pretrained 3DCNN	6	0.027344
XGBOOST	cPGS Pretrained 3DCNN with cPGS	10	0.083984	XGBOOST	cPGS Pretrained 3DCNN with cPGS	26	0.921875	XGBOOST	cPGS Pretrained 3DCNN with cPGS	5	0.019531
scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	5	0.12819	scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	9	0.109745	scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	4	0.028402
scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	10	0.262033	scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	8.5	0.182893	scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	2	0.005859

	with cPGS				with cPGS				with cPGS		
cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	11	0.326989	cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	12.5	0.440648	cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	13	0.260393

Supplementary Table 16– 2y MDD Prediction Model Comparison_Modalities (DEP PGS)

ACC				AUROC				F1			
Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
DEP PGS MLP	DEP PGS Pretrained 3DCNN	0	0.001953	DEP PGS MLP	DEP PGS Pretrained 3DCNN	0	0.001953	DEP PGS MLP	DEP PGS Pretrained 3DCNN	0	0.001953
DEP PGS MLP	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	DEP PGS MLP	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	DEP PGS MLP	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
from-scratch 3DCNN	DEP PGS Pretrained 3DCNN	11	0.105469	from-scratch 3DCNN	DEP PGS Pretrained 3DCNN	10	0.083984	from-scratch 3DCNN	DEP PGS Pretrained 3DCNN	0	0.001953
from-scratch 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	1	0.0458	from-scratch 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	from-scratch 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	8.5	0.182893	from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	4	0.013672	from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	0	0.001953
from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	6	0.092491	from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	1	0.003906	from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	18.5	0.633794	DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	1	0.003906

Supplementary Table 17– 2y MDD Prediction Model Comparison_Modalities (cPGS)

ACC				AUROC				F1			
Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
cPGS MLP	cPGS Pretrained 3DCNN	0	0.001953	cPGS MLP	cPGS Pretrained 3DCNN	0	0.001953	cPGS MLP	cPGS Pretrained 3DCNN	0	0.001953
cPGS MLP	cPGS Pretrained 3DCNN with cPGS	0	0.001953	cPGS MLP	cPGS Pretrained 3DCNN with cPGS	0	0.001953	cPGS MLP	cPGS Pretrained 3DCNN with cPGS	0	0.001953

from-scratch 3DCNN	cPGS Pretrained 3DCNN	3	0.062509	from-scratch 3DCNN	cPGS Pretrained 3DCNN	4	0.013672	from-scratch 3DCNN	cPGS Pretrained 3DCNN	0	0.001953
from-scratch 3DCNN	cPGS Pretrained 3DCNN with cPGS	2.5	0.017518	from-scratch 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953	from-scratch 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953
from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	3	0.020433	from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	24	0.769531	from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	8	0.048828
from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	8.5	0.048828	from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	0	0.001953	from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	0	0.001953
cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953	cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953	cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953

Supplementary Table 18 – 2y MDD Prediction Model Comparison_Machine Learning (DEP PGS)

ACC				AUROC				F1			
Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
SVC_linear	DEP PGS Pretrained 3DCNN	0	0.001953	SVC_linear	DEP PGS Pretrained 3DCNN	0	0.001953	SVC_linear	DEP PGS Pretrained 3DCNN	0	0.001953
SVC_linear	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	SVC_linear	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	SVC_linear	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
SVC_rbf	DEP PGS Pretrained 3DCNN	0	0.001953	SVC_rbf	DEP PGS Pretrained 3DCNN	0	0.001953	SVC_rbf	DEP PGS Pretrained 3DCNN	0	0.001953
SVC_rbf	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	SVC_rbf	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	SVC_rbf	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
XGBOOST	DEP PGS Pretrained 3DCNN	0	0.001953	XGBOOST	DEP PGS Pretrained 3DCNN	0	0.001953	XGBOOST	DEP PGS Pretrained 3DCNN	0	0.001953
XGBOOST	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	XGBOOST	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	XGBOOST	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	8.5	0.182893	scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	4	0.013672	scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	0	0.001953
scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	6	0.092491	scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	1	0.003906	scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN	18.5	0.633794	DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN	0	0.001953	DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN	1	0.003906

	with DEP PGS				with DEP PGS				with DEP PGS		

Supplementary Table 19 – 2y MDD Prediction Model Comparison_Machine Learning (cPGS)

ACC				AUROC				F1			
Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
SVC_linear	cPGS Pretrained 3DCNN	0	0.001953	SVC_linear	cPGS Pretrained 3DCNN	0	0.001953	SVC_linear	cPGS Pretrained 3DCNN	0	0.001953
SVC_linear	cPGS Pretrained 3DCNN with cPGS	0	0.001953	SVC_linear	cPGS Pretrained 3DCNN with cPGS	0	0.001953	SVC_linear	cPGS Pretrained 3DCNN with cPGS	0	0.001953
SVC_rbf	cPGS Pretrained 3DCNN	0	0.001953	SVC_rbf	cPGS Pretrained 3DCNN	0	0.001953	SVC_rbf	cPGS Pretrained 3DCNN	0	0.001953
SVC_rbf	cPGS Pretrained 3DCNN with cPGS	0	0.001953	SVC_rbf	cPGS Pretrained 3DCNN with cPGS	0	0.001953	SVC_rbf	cPGS Pretrained 3DCNN with cPGS	0	0.001953
XGBOOST	cPGS Pretrained 3DCNN	0	0.001953	XGBOOST	cPGS Pretrained 3DCNN	0	0.001953	XGBOOST	cPGS Pretrained 3DCNN	0	0.001953
XGBOOST	cPGS Pretrained 3DCNN with cPGS	0	0.001953	XGBOOST	cPGS Pretrained 3DCNN with cPGS	0	0.001953	XGBOOST	cPGS Pretrained 3DCNN with cPGS	0	0.001953
scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	3	0.020433	scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	24	0.769531	scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	8	0.048828
scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	8.5	0.048828	scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	0	0.001953	scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	0	0.001953
cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953	cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953	cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953

Supplementary Table 20– 2y MDD + SIA Prediction Model Comparison_Modalities (DEP PGS)

ACC				AUROC				F1			
Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
DEP PGS MLP	DEP PGS Pretrained 3DCNN	1	0.003906	DEP PGS MLP	DEP PGS Pretrained 3DCNN	1	0.003906	DEP PGS MLP	DEP PGS Pretrained 3DCNN	0	0.001953
DEP PGS MLP	DEP PGS Pretrained 3DCNN with DEP PGS	1	0.003906	DEP PGS MLP	DEP PGS Pretrained 3DCNN with DEP PGS	1	0.003906	DEP PGS MLP	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953

from-scratch 3DCNN	DEP PGS Pretrained 3DCNN	7.5	0.075304	from-scratch 3DCNN	DEP PGS Pretrained 3DCNN	8	0.048828	from-scratch 3DCNN	DEP PGS Pretrained 3DCNN	0	0.001953
from-scratch 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	from-scratch 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	from-scratch 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	0	0.007474	from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	0	0.001953	from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	2	0.005859
from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	9	0.108497	from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953

Supplementary Table 21– 2y MDD + SIA Prediction Model Comparison_Modalities (cPGS)

ACC				AUROC				F1			
Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
cPGS MLP	cPGS Pretrained 3DCNN	9	0.064453	cPGS MLP	cPGS Pretrained 3DCNN	3	0.009766	cPGS MLP	cPGS Pretrained 3DCNN	0	0.001953
cPGS MLP	cPGS Pretrained 3DCNN with cPGS	0	0.001953	cPGS MLP	cPGS Pretrained 3DCNN with cPGS	14	0.193359	cPGS MLP	cPGS Pretrained 3DCNN with cPGS	0	0.001953
from-scratch 3DCNN	cPGS Pretrained 3DCNN	2	0.005859	from-scratch 3DCNN	cPGS Pretrained 3DCNN	0	0.001953	from-scratch 3DCNN	cPGS Pretrained 3DCNN	0	0.001953
from-scratch 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953	from-scratch 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953	from-scratch 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953
from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	2.5	0.009766	from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	9	0.064453	from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	0	0.001953
from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	5.5	0.019531	from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	0	0.001953	from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	0	0.001953
cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953	cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953	cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953

Supplementary Table 22 – 2y MDD +SIA prediction model Comparison_Machine Learning (DEP PGS)

ACC				AUROC				F1			

Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
SVC_linear	DEP PGS Pretrained 3DCNN	0	0.001953	SVC_linear	SVC_linear	DEP PGS Pretrained 3DCNN	5	0.019531	DEP PGS Pretrained 3DCNN	0	0.001953
SVC_linear	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	SVC_linear	SVC_linear	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
SVC_rbf	DEP PGS Pretrained 3DCNN	0	0.001953	SVC_rbf	SVC_rbf	DEP PGS Pretrained 3DCNN	0	0.001953	DEP PGS Pretrained 3DCNN	0	0.001953
SVC_rbf	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	SVC_rbf	SVC_rbf	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
XGBOOST	DEP PGS Pretrained 3DCNN	0	0.001953	XGBOOST	XGBOOST	DEP PGS Pretrained 3DCNN	1	0.003906	DEP PGS Pretrained 3DCNN	0	0.001953
XGBOOST	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	XGBOOST	XGBOOST	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	0	0.007474	scratch 3DCNN with cPGS	scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	0	0.001953	DEP PGS Pretrained 3DCNN	2	0.005859
scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	9	0.108497	scratch 3DCNN with cPGS	scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	cPGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953

Supplementary Table 23 – 2y MDD +SIA prediction model Comparison_Machine Learning (cPGS)

ACC				AUROC				F1			
Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
SVC_linear	cPGS Pretrained 3DCNN	0	0.001953	SVC_linear	cPGS Pretrained 3DCNN	3	0.009766	SVC_linear	cPGS Pretrained 3DCNN	0	0.001953
SVC_linear	cPGS Pretrained 3DCNN with cPGS	0	0.001953	SVC_linear	cPGS Pretrained 3DCNN with cPGS	0	0.001953	SVC_linear	cPGS Pretrained 3DCNN with cPGS	0	0.001953
SVC_rbf	cPGS Pretrained 3DCNN	0	0.001953	SVC_rbf	cPGS Pretrained 3DCNN	0	0.001953	SVC_rbf	cPGS Pretrained 3DCNN	0	0.001953
SVC_rbf	cPGS Pretrained 3DCNN	0	0.001953	SVC_rbf	cPGS Pretrained 3DCNN	0	0.001953	SVC_rbf	cPGS Pretrained 3DCNN	0	0.001953

	with cPGS				with cPGS				with cPGS		
XGBOOST	cPGS Pretrained 3DCNN	1	0.003906	XGBOOST	cPGS Pretrained 3DCNN	0	0.001953	XGBOOST	cPGS Pretrained 3DCNN	0	0.001953
XGBOOST	cPGS Pretrained 3DCNN with cPGS	0	0.001953	XGBOOST	cPGS Pretrained 3DCNN with cPGS	0	0.001953	XGBOOST	cPGS Pretrained 3DCNN with cPGS	0	0.001953
scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	2.5	0.009766	scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	9	0.064453	scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	0	0.001953
scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	5.5	0.019531	scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	0	0.001953	scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	0	0.001953
cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953	cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953	cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953

Supplementary Table 24 – 2y MDD + SA Prediction Model Comparison_Modalities (DEP PGS)

ACC				AUROC				F1			
Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
DEP PGS MLP	DEP PGS Pretrained 3DCNN	5.5	0.027344	DEP PGS MLP	DEP PGS Pretrained 3DCNN	0	0.001953	DEP PGS MLP	DEP PGS Pretrained 3DCNN	3	0.009766
DEP PGS MLP	DEP PGS Pretrained 3DCNN with DEP PGS	12.5	0.235314	DEP PGS MLP	DEP PGS Pretrained 3DCNN with DEP PGS	13	0.160156	DEP PGS MLP	DEP PGS Pretrained 3DCNN with DEP PGS	15	0.232422
from-scratch 3DCNN	DEP PGS Pretrained 3DCNN	9.5	0.122546	from-scratch 3DCNN	DEP PGS Pretrained 3DCNN	10	0.083984	from-scratch 3DCNN	DEP PGS Pretrained 3DCNN	0	0.001953
from-scratch 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	1	0.010724	from-scratch 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	from-scratch 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	0	0.007421	from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	0	0.001953	from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	0	0.007632
from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	6	0.340085	from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	14	0.574464	from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	2	0.005859
DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	4	0.04967	DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953

Supplementary Table 25 – 2y MDD + SA Prediction Model Comparison_Modalities (cPGS)

ACC				AUROC				F1			
Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
cPGS MLP	cPGS Pretrained 3DCNN	13	0.482765	cPGS MLP	cPGS Pretrained 3DCNN	7	0.037109	cPGS MLP	cPGS Pretrained 3DCNN	1	0.003906
cPGS MLP	cPGS Pretrained 3DCNN with cPGS	20	0.767097	cPGS MLP	cPGS Pretrained 3DCNN with cPGS	26	0.921875	cPGS MLP	cPGS Pretrained 3DCNN with cPGS	0	0.001953
from-scratch 3DCNN	cPGS Pretrained 3DCNN	1	0.010724	from-scratch 3DCNN	cPGS Pretrained 3DCNN	27.5	1	from-scratch 3DCNN	cPGS Pretrained 3DCNN	0	0.001953
from-scratch 3DCNN	cPGS Pretrained 3DCNN with cPGS	1	0.010586	from-scratch 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953	from-scratch 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953
from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	13.5	0.526066	from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	19	0.431641	from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	9.5	0.123205
from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	6.5	0.203272	from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	5	0.019531	from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	0	0.007686
cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	3	0.113846	cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953	cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953

Supplementary Table 26 – 2y MDD +SA Prediction Model Comparison_Machine Learning (DEP PGS)

ACC				AUROC				F1			
Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
SVC_linear	DEP PGS Pretrained 3DCNN	3	0.009766	SVC_linear	DEP PGS Pretrained 3DCNN	23	0.695313	SVC_linear	DEP PGS Pretrained 3DCNN	0	0.001953
SVC_linear	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	SVC_linear	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	SVC_linear	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
SVC_rbf	DEP PGS Pretrained 3DCNN	9	0.109121	SVC_rbf	DEP PGS Pretrained 3DCNN	6.5	0.027344	SVC_rbf	DEP PGS Pretrained 3DCNN	0	0.001953
SVC_rbf	DEP PGS Pretrained 3DCNN with DEP PGS	2	0.014978	SVC_rbf	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	SVC_rbf	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
XGBOOST	DEP PGS Pretrained 3DCNN	15	0.232422	XGBOOST	DEP PGS Pretrained 3DCNN	22	0.952765	XGBOOST	DEP PGS Pretrained 3DCNN	18	0.375
XGBOOST	DEP PGS Pretrained 3DCNN	4	0.013672	XGBOOST	DEP PGS Pretrained 3DCNN	0	0.001953	XGBOOST	DEP PGS Pretrained 3DCNN	3	0.009766

	with DEP PGS				with DEP PGS				with DEP PGS		
scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	0	0.007421	scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	0	0.001953	scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	0	0.007632
scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	6	0.340085	scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	14	0.574464	scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	2	0.005859
DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	4	0.04967	DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953

Supplementary Table 27 – 2y MDD +SA Prediction Model Comparison_Machine Learning (cPGS)

ACC				AUROC				F1			
Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
SVC_linear	cPGS Pretrained 3DCNN	0	0.007579	SVC_linear	cPGS Pretrained 3DCNN	13	0.160156	SVC_linear	cPGS Pretrained 3DCNN	0	0.001953
SVC_linear	cPGS Pretrained 3DCNN with cPGS	0	0.007632	SVC_linear	cPGS Pretrained 3DCNN with cPGS	0	0.001953	SVC_linear	cPGS Pretrained 3DCNN with cPGS	0	0.001953
SVC_rbf	cPGS Pretrained 3DCNN	1	0.003906	SVC_rbf	cPGS Pretrained 3DCNN	0	0.001953	SVC_rbf	cPGS Pretrained 3DCNN	0	0.001953
SVC_rbf	cPGS Pretrained 3DCNN with cPGS	1	0.003906	SVC_rbf	cPGS Pretrained 3DCNN with cPGS	0	0.001953	SVC_rbf	cPGS Pretrained 3DCNN with cPGS	0	0.001953
XGBOOST	cPGS Pretrained 3DCNN	9	0.064453	XGBOOST	cPGS Pretrained 3DCNN	13	0.160156	XGBOOST	cPGS Pretrained 3DCNN	2	0.005859
XGBOOST	cPGS Pretrained 3DCNN with cPGS	9	0.064453	XGBOOST	cPGS Pretrained 3DCNN with cPGS	2	0.005859	XGBOOST	cPGS Pretrained 3DCNN with cPGS	1	0.003906
scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	13.5	0.526066	scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	19	0.431641	scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	9.5	0.123205
scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	6.5	0.203272	scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	5	0.019531	scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	0	0.007686
cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	3	0.113846	cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953	cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953

Supplementary Table 28 – Control Analysis Performance Comparison

ACC				AUROC				F1			
Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
ABCD family history pretrained 3DCNN	ABCD totprob pretrained 3DCNN	17	0.322266	ABCD family history pretrained 3DCNN	ABCD totprob pretrained 3DCNN	13	0.160156	ABCD family history pretrained 3DCNN	ABCD totprob pretrained 3DCNN	22	0.625
ABCD family history pretrained 3DCNN	ABCD intelligence pretrained 3DCNN	12	0.213524	ABCD family history pretrained 3DCNN	ABCD intelligence pretrained 3DCNN	17	0.322266	ABCD family history pretrained 3DCNN	ABCD intelligence pretrained 3DCNN	6	0.027344
ABCD family history pretrained 3DCNN	ABCD tck cPGS pretrained 3DCNN	12	0.735317	ABCD family history pretrained 3DCNN	ABCD tck cPGS pretrained 3DCNN	26	0.921875	ABCD family history pretrained 3DCNN	ABCD tck cPGS pretrained 3DCNN	20	0.492188
ABCD family history pretrained 3DCNN	ABCD_DEP_PGS pretrained 3DCNN	5	0.038152	ABCD family history pretrained 3DCNN	ABCD_DEP_PGS pretrained 3DCNN	3	0.009766	ABCD family history pretrained 3DCNN	ABCD_DEP_PGS pretrained 3DCNN	8	0.048828
ABCD family history pretrained 3DCNN	ABCD_cPGS pretrained 3DCNN	4	0.013672	ABCD family history pretrained 3DCNN	ABCD_cPGS pretrained 3DCNN	0	0.001953	ABCD family history pretrained 3DCNN	ABCD_cPGS pretrained 3DCNN	7	0.037109
ABCD totprob pretrained 3DCNN	ABCD intelligence pretrained 3DCNN	11	0.173071	ABCD totprob pretrained 3DCNN	ABCD intelligence pretrained 3DCNN	18	0.375	ABCD totprob pretrained 3DCNN	ABCD intelligence pretrained 3DCNN	17	0.322266
ABCD totprob pretrained 3DCNN	ABCD tck cPGS pretrained 3DCNN	12	0.130859	ABCD totprob pretrained 3DCNN	ABCD tck cPGS pretrained 3DCNN	17	0.322266	ABCD totprob pretrained 3DCNN	ABCD tck cPGS pretrained 3DCNN	9	0.109745
ABCD totprob pretrained 3DCNN	ABCD_DEP_PGS pretrained 3DCNN	23	0.695313	ABCD totprob pretrained 3DCNN	ABCD_DEP_PGS pretrained 3DCNN	20	0.492188	ABCD totprob pretrained 3DCNN	ABCD_DEP_PGS pretrained 3DCNN	12	0.130859
ABCD totprob pretrained 3DCNN	ABCD_cPGS pretrained 3DCNN	23	0.695313	ABCD totprob pretrained 3DCNN	ABCD_cPGS pretrained 3DCNN	25	0.845703	ABCD totprob pretrained 3DCNN	ABCD_cPGS pretrained 3DCNN	9	0.064453
ABCD intelligence pretrained 3DCNN	ABCD tck cPGS pretrained 3DCNN	20	0.492188	ABCD intelligence pretrained 3DCNN	ABCD tck cPGS pretrained 3DCNN	18	0.375	ABCD intelligence pretrained 3DCNN	ABCD tck cPGS pretrained 3DCNN	8	0.048828
ABCD intelligence pretrained 3DCNN	ABCD_DEP_PGS pretrained 3DCNN	1	0.010862	ABCD intelligence pretrained 3DCNN	ABCD_DEP_PGS pretrained 3DCNN	14	0.193359	ABCD intelligence pretrained 3DCNN	ABCD_DEP_PGS pretrained 3DCNN	13	0.160156
ABCD intelligence pretrained 3DCNN	ABCD_cPGS pretrained 3DCNN	5	0.019531	ABCD intelligence pretrained 3DCNN	ABCD_cPGS pretrained 3DCNN	7	0.037109	ABCD intelligence pretrained 3DCNN	ABCD_cPGS pretrained 3DCNN	12	0.213524
ABCD tck cPGS pretrained 3DCNN	ABCD_DEP_PGS pretrained 3DCNN	9	0.109745	ABCD tck cPGS pretrained 3DCNN	ABCD_DEP_PGS pretrained 3DCNN	14	0.193359	ABCD tck cPGS pretrained 3DCNN	ABCD_DEP_PGS pretrained 3DCNN	4	0.028402

ABCD tck cPGS pretrained 3DCNN	ABCD_cPGS pretrained 3DCNN	13	0.160156	ABCD tck cPGS pretrained 3DCNN	ABCD_cPGS pretrained 3DCNN	9	0.064453	ABCD tck cPGS pretrained 3DCNN	ABCD_cPGS pretrained 3DCNN	5	0.019531
ABCD_DEP_PGS pretrained 3DCNN	ABCD_cPGS pretrained 3DCNN	11.5	0.672052	ABCD_DEP_PGS pretrained 3DCNN	ABCD_cPGS pretrained 3DCNN	0	0.011719	ABCD_DEP_PGS pretrained 3DCNN	ABCD_cPGS pretrained 3DCNN	27	1

Supplementary Table 29 –2y MDD without Suicidal Behavior Prediction Model Comparison_Modalities (DEP PGS)

ACC				AUROC				F1			
Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
DEP PGS MLP	DEP PGS Pretrained 3DCNN	0	0.001953	DEP PGS MLP	DEP PGS Pretrained 3DCNN	0	0.001953	DEP PGS MLP	DEP PGS Pretrained 3DCNN	0	0.001953
DEP PGS MLP	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	DEP PGS MLP	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	DEP PGS MLP	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
from-scratch 3DCNN	DEP PGS Pretrained 3DCNN	3	0.035015	from-scratch 3DCNN	DEP PGS Pretrained 3DCNN	22.5	0.625	from-scratch 3DCNN	DEP PGS Pretrained 3DCNN	0	0.001953
from-scratch 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	1	0.003906	from-scratch 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	16	0.441268	from-scratch 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	0	0.027281	from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	20	0.492188	from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	0	0.001953
from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	1	0.003906	from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	7	0.037109	from-scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	1	0.003906	DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	18	0.375	DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	1	0.003906

Supplementary Table 30 –2y MDD without Suicidal Behavior Prediction Model Comparison_Modalities (cPGS)

ACC				AUROC				F1			
Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
cPGS MLP	cPGS Pretrained 3DCNN	0	0.001953	cPGS MLP	cPGS Pretrained 3DCNN	0	0.001953	cPGS MLP	cPGS Pretrained 3DCNN	0	0.001953
cPGS MLP	cPGS Pretrained	0	0.001953	cPGS MLP	cPGS Pretrained	0	0.001953	cPGS MLP	cPGS Pretrained	0	0.001953

	3DCNN with cPGS				3DCNN with cPGS				3DCNN with cPGS		
from-scratch 3DCNN	cPGS Pretrained 3DCNN	0	0.001953	from-scratch 3DCNN	cPGS Pretrained 3DCNN	21	0.556641	from-scratch 3DCNN	cPGS Pretrained 3DCNN	26	0.921875
from-scratch 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953	from-scratch 3DCNN	cPGS Pretrained 3DCNN with cPGS	3.5	0.009766	from-scratch 3DCNN	cPGS Pretrained 3DCNN with cPGS	18	0.593955
from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	15	0.674424	from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	14.5	0.623212	from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	20	0.492188
from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	0	0.001953	from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	0	0.007526	from-scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	23.5	0.695313
cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953	cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	3	0.020767	cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	18	0.375

Supplementary Table 31 – 2y MDD without Suicidal Behavior Prediction Model Comparison_Machine Learning (DEP PGS)

ACC				AUROC				F1			
Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
SVC_linear	DEP PGS Pretrained 3DCNN	0	0.001953	SVC_linear	DEP PGS Pretrained 3DCNN	0	0.001953	SVC_linear	DEP PGS Pretrained 3DCNN	0	0.001953
SVC_linear	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	SVC_linear	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	SVC_linear	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
SVC_rbf	DEP PGS Pretrained 3DCNN	0	0.001953	SVC_rbf	DEP PGS Pretrained 3DCNN	0	0.001953	SVC_rbf	DEP PGS Pretrained 3DCNN	0	0.001953
SVC_rbf	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	SVC_rbf	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	SVC_rbf	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
XGBOOST	DEP PGS Pretrained 3DCNN	0	0.001953	XGBOOST	DEP PGS Pretrained 3DCNN	0	0.001953	XGBOOST	DEP PGS Pretrained 3DCNN	0	0.001953
XGBOOST	DEP PGS Pretrained 3DCNN with DEP PGS	7	0.037109	XGBOOST	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953	XGBOOST	DEP PGS Pretrained 3DCNN with DEP PGS	0	0.001953
scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	0	0.027281	scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	20	0.492188	scratch 3DCNN with DEP PGS	DEP PGS Pretrained 3DCNN	0	0.001953
scratch 3DCNN	DEP PGS Pretrained 3DCNN	1	0.003906	scratch 3DCNN with DEP PGS	DEP PGS Pretrained	12	0.130859	scratch 3DCNN	DEP PGS Pretrained 3DCNN	0	0.001953

with DEP PGS	with DEP PGS				3DCNN with DEP PGS			with DEP PGS	with DEP PGS		
DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	1	0.003906	DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	13	0.160156	DEP PGS Pretrained 3DCNN	DEP PGS Pretrained 3DCNN with DEP PGS	1	0.003906

Supplementary Table 32 – 2y MDD without Suicidal Behavior Prediction Model Comparison_Machine Learning (cPGS)

ACC				AUROC				F1			
Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
SVC_linear	cPGS Pretrained 3DCNN	0	0.001953	SVC_linear	cPGS Pretrained 3DCNN	0	0.001953	SVC_linear	cPGS Pretrained 3DCNN	0	0.001953
SVC_linear	cPGS Pretrained 3DCNN with cPGS	0	0.001953	SVC_linear	cPGS Pretrained 3DCNN with cPGS	0	0.001953	SVC_linear	cPGS Pretrained 3DCNN with cPGS	0	0.001953
SVC_rbf	cPGS Pretrained 3DCNN	0	0.001953	SVC_rbf	cPGS Pretrained 3DCNN	1	0.003906	SVC_rbf	cPGS Pretrained 3DCNN	0	0.001953
SVC_rbf	cPGS Pretrained 3DCNN with cPGS	0	0.001953	SVC_rbf	cPGS Pretrained 3DCNN with cPGS	1	0.003906	SVC_rbf	cPGS Pretrained 3DCNN with cPGS	0	0.001953
XGBOOST	cPGS Pretrained 3DCNN	1	0.003906	XGBOOST	cPGS Pretrained 3DCNN	0	0.001953	XGBOOST	cPGS Pretrained 3DCNN	0	0.001953
XGBOOST	cPGS Pretrained 3DCNN with cPGS	0	0.001953	XGBOOST	cPGS Pretrained 3DCNN with cPGS	0	0.001953	XGBOOST	cPGS Pretrained 3DCNN with cPGS	0	0.001953
scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	14.5	0.623212	scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	20	0.492188	scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN	15	0.674424
scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	0	0.007526	scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	23.5	0.695313	scratch 3DCNN with cPGS	cPGS Pretrained 3DCNN with cPGS	0	0.001953
cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	3	0.020767	cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	18	0.375	cPGS Pretrained 3DCNN	cPGS Pretrained 3DCNN with cPGS	0	0.001953

Supplementary Table 33 – Korean dataset Depression Classification model Comparison

ACC				AUROC				F1			
Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value	Model 1	Model 2	Statistic	P-Value
from_scratch 3DCNN	ABCD_DEP_PGS pretrained 3DCNN	14	1	from_scratch 3DCNN	ABCD_DEP_PGS pretrained 3DCNN	6	0.049602	from_scratch 3DCNN	ABCD_DEP_PGS pretrained 3DCNN	21	0.858955

from_scratch 3DCNN	ABCD_cPGS pretrained 3DCNN	8.5	0.094264	from_scratch 3DCNN	ABCD_cPGS pretrained 3DCNN	2	0.014802	from_scratch 3DCNN	ABCD_cPGS pretrained 3DCNN	16	0.275391
ABCD_from_scratch MDD pretrained 3DCNN	ABCD_DEP_PGS pretrained 3DCNN	26	0.921875	ABCD_from_scratch MDD pretrained 3DCNN	ABCD_DEP_PGS pretrained 3DCNN	10	0.083984	ABCD_from_scratch MDD pretrained 3DCNN	ABCD_DEP_PGS pretrained 3DCNN	16	0.275391

References

1. Casey, B.J., et al., *The adolescent brain cognitive development (ABCD) study: imaging acquisition across 21 sites.* Developmental cognitive neuroscience, 2018. **32**: p. 43-54.
2. Hagler Jr, D.J., et al., *Image processing and analysis methods for the Adolescent Brain Cognitive Development Study.* Neuroimage, 2019. **202**: p. 116091.
3. Tournier, J.-D., et al., *MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation.* Neuroimage, 2019. **202**: p. 116137.
4. Mori, S. and J. Zhang, *Principles of diffusion tensor imaging and its applications to basic neuroscience research.* Neuron, 2006. **51**(5): p. 527-539.
5. Mukherjee, P., et al., *Normal brain maturation during childhood: developmental trends characterized with diffusion-tensor MR imaging.* Radiology, 2001. **221**(2): p. 349-358.
6. Tournier, J.-D., F. Calamante, and A. Connelly, *Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution.* Neuroimage, 2007. **35**(4): p. 1459-1472.
7. Smith, R.E., et al., *SIFT: Spherical-deconvolution informed filtering of tractograms.* Neuroimage, 2013. **67**: p. 298-312.
8. Hayashi, S., et al., *brainlife. io: a decentralized and open-source cloud platform to support neuroscience research.* Nature methods, 2024. **21**(5): p. 809-813.
9. Kline, A. and Y. Luo. *PsmPy: a package for retrospective cohort matching in Python*. in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. 2022. IEEE.
10. Consortium, G.P., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68.
11. Conomos, M.P., M.B. Miller, and T.A. Thornton, *Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness.* Genetic epidemiology, 2015. **39**(4): p. 276-293.
12. Conomos, M.P., et al., *Model-free estimation of recent genetic relatedness.* The American Journal of Human Genetics, 2016. **98**(1): p. 127-148.
13. Joo, Y.Y., et al., *Polygenic architecture of brain structure and function, behaviors, and psychopathologies in children.* bioRxiv, 2024: p. 2024.05.22.595444.
14. Shen, X., et al., *Subcortical volume and white matter integrity abnormalities in major depressive disorder: findings from UK Biobank imaging data.* Scientific reports, 2017. **7**(1): p. 5547.