AI-Powered Brain Mapping for Psychiatric Diagnostics

A Vision Transformer Framework Using 3D T1-Weighted MRI

Project Summary / Abstract

Psychiatric disorders such as schizophrenia (SCZ), bipolar disorder (BD), obsessive-compulsive disorder (OCD), and major depressive disorder (MDD) remain difficult to diagnose and monitor objectively, due to their heterogeneous presentations and lack of reliable biomarkers. Structural brain changes in these disorders are well-documented, yet not leveraged effectively for individual-level diagnosis or symptom tracking.

We propose to develop and validate a precision neuroimaging framework that uses only 3D T1-weighted (T1w) structural MRI, the most accessible and widely used modality in clinical neuroimaging. Our approach leverages 3D Vision Transformer (ViT) architectures to perform (1) diagnostic classification of SCZ, BD, OCD, MDD, and healthy controls, (2) prediction of symptom severity from structural scans, and (3) visualization of disorder-specific spatial patterns using explainable AI (XAI) techniques. This project will generate AI-based neuroimaging biomarkers that are clinically scalable, interpretable, and applicable at the individual level, bringing psychiatry closer to personalized medicine.

Specific Aims

Aim 1

Develop and optimize a 3D vision transformer model for structural MRI-based classification of SCZ, BD, OCD, and MDD.
We will build a vision transformer model that classifies patients into psychiatric diagnostic categories using only 3D T1-weighted MRI. We will train and evaluate the model using multi-site, multi-diagnosis structural MRI datasets. Comparisons will be made to standard 3D CNNs.
Hypothesis: The transformer-based model will achieve higher accuracy, generalizability, and interpretability compared to traditional CNNs.

Aim 2

Predict individual-level symptom severity scores from T1-weighted MRI using learned AI features.
We will use model-derived embeddings and regression heads to predict standardized clinical scores such as PANSS (SCZ), MADRS (MDD), YBOCS (OCD), and HAM-D. We will also examine the relationship between AI-predicted severity and demographic or cognitive metrics using canonical correlation analysis (CCA).

Hypothesis: Neuroimaging-derived AI scores will show significant correlations with symptom scales, providing an objective tool to measure psychiatric symptom severity.

Aim 3

Generate spatial brain fingerprints of psychiatric disorders using explainable AI.
We will apply explainability techniques to identify brain regions contributing to diagnostic classification across disorders. Group-level attention maps will be used to create "brain fingerprints" that characterize each disorder's unique and shared anatomical features.
Hypothesis: Each disorder will show distinct and interpretable structural features localized to cortical and subcortical regions.

Research Strategy

Significance

Current psychiatric diagnosis relies heavily on self-report and clinician interpretation. Neuroimaging has the potential to provide objective biomarkers, but has yet to reach individualized clinical utility. Our proposed T1w MRI-based AI framework could:
- Provide an accessible and cost-effective diagnostic tool,
- Reveal latent structural signatures of psychiatric disorders,
- Enable continuous tracking of symptom severity, and
- Guide targeted treatment by identifying disorder-specific anatomical disruptions.

Innovation

- First use of 3D Vision Transformers trained on T1w MRI for multi-disorder psychiatric classification.
- Symptom-level prediction from structural neuroimaging using deep feature embeddings.
- Use of explainable AI to generate spatial disease-specific fingerprints.
- Exclusive use of T1-weighted MRI, making the model compatible with routine clinical imaging and large biobanks.

Approach

We propose a scalable, non-invasive AI framework using only 3D T1-weighted MRI—the most accessible structural neuroimaging modality. Our approach includes data preprocessing, model development, severity prediction, and spatial explainability.

Data Acquisition & Preprocessing:
- Use open-access datasets (COBRE, MCIC, LA5c, BSNIP, OASIS-3, UK Biobank)
- Preprocess with FreeSurfer and ANTs (bias correction, registration to MNI, resampling to $1mm^3$)

Model Development:
- Develop and optimize Swin Transformer 3D, ViT-UNet, or hybrid CNN-ViT architectures
- Input: 3D T1w MRI; Output: diagnosis + continuous severity score
- Training: Cross-entropy + MSE loss; augmentation via TorchIO; cross-validation and external testing

Symptom Prediction:
- Use transformer encoder embeddings
- Predict PANSS, MADRS, YBOCS, HAM-D
- Correlate predictions with clinical and demographic data using regression and CCA

Explainable AI:
- Use attention rollout, SHAP, Integrated Gradients
- Create spatial “brain fingerprints” by diagnosis

Timeline

Year 1: Dataset curation, T1w preprocessing, SCZ vs Control replication using ViT
Year 2: Full multi-disorder classification model and symptom score regression
Year 3: XAI spatial fingerprint generation, statistical validation, publication/dissemination

Conclusion

This project lays the foundation for scalable, non-invasive, precision neuropsychiatry using the most accessible MRI modality—T1-weighted structural imaging. Our use of AI, specifically 3D vision transformers, will unlock rich structural signatures of psychiatric disease and enable individualized assessments that are interpretable, accurate, and widely deployable. The result is a step toward transforming psychiatry from symptom-based diagnosis to data-driven precision medicine.