Early Diagnosis of Schizophrenia Through an Interpretable CNN–Transformer Model of fMRI Connectivity

Lyakhov, Michael (School: John L. Miller Great Neck North High School)

Schizophrenia affects 24 million people worldwide. A critical barrier to treatment is anosognosia, the lack of illness awareness that develops in later-stage SCZ and causes patients to refuse treatment. Early diagnosis before psychosis onset is therefore essential, yet chemical biomarkers lack specificity and cognitive assessments fail at the prodromal stage. This study presents an interpretable CNN-Transformer deep learning model that classifies schizophrenia and predicts treatment trajectory from resting-state fMRI functional connectivity. Preprocessing and feature extraction were performed on data from 1,684 SCZ patients and 1,684 healthy controls. A 3D CNN Spatial Encoder extracts spatial embeddings from fMRI time windows, which a multi-head Transformer Temporal Encoder processes sequentially. A Prognostic Head then classifies subjects into three clinically actionable categories: stable control, treatment-responsive SCZ, and progressive psychosis/treatment resistance, a stratification previously impossible without longitudinal clinical observation. The model achieved 91.3% accuracy, AUC 0.92, 86.4% sensitivity, and 92.9% specificity. SHAP analysis confirmed ACC-DLPFC fronto-cingulate dysconnectivity as the dominant biomarker, providing neurobiologically grounded interpretability. The pipeline is published as an open-source Python library on PyPI for direct clinical use. This work establishes that interpretable deep learning on fMRI connectivity can enable early SCZ diagnosis and treatment stratification before anosognosia develops, directly improving long-term patient outcomes.

Awards Won:

Second Award of $2,400