[BIOINFO 2025]

Interactive Visualization and Harmonization of Multicenter Brain Imaging Cohorts

Byung Soo Park[1], Myeongji Cho[1], Hye Ryeong Nam[1], Jeong Won Hong[1] and Sang Cheol Kim[1,*]

[1]*Division of Healthcare and Artificial Intelligence, Department of Precision Medicine, National Institute of Health, Korea Disease Control and Prevention Agency, Cheong-Ju, Republic of Korea.*
**Corresponding author: sckim.knih@korea.kr*

Brain imaging plays a key role in the study of neurodegenerative disease, enabling researchers to associate structural brain features with clinical and epidemiological factors. While most existing visualization systems focus on individual cases or single regions, large multicenter cohorts require interactive platforms that integrate multimodal data for comparative and exploratory analysis.

We introduce an interactive web-based tool, developed in Python (Streamlit, Plotly), that enables the joint exploration of brain imaging, clinical, and epidemiological data collected across centers. The system allows flexible selection of categorical and continuous variables, visualized through scatter, box, and violin plots, with built-in statistical tests (pairwise t-tests, ANOVA) conditioned on user-defined strata. To capture cohort-level structure, dimensionality reduction (PCA) is provided for pattern discovery and for detecting site- or batch-related effects.

The tool features a configurable preprocessing pipeline, including log transformation, standard scaling, intracranial-volume (ICV) normalization, optional outlier trimming (0.1th–99.9th percentiles), and ComBat-based harmonization with age/sex covariates to improve cross-site comparability. A 'prominent regions' mode links statistical summaries with anatomical thumbnails to streamline region-focused exploration. For reproducible and efficient sessions, uploaded CSVs are converted to compressed feather files and cached, enabling seamless re-loading and side-by-side comparison of raw versus harmonized data. Together, these capabilities provide a practical environment for cohort-scale, multimodal interrogation of brain imaging data and support the generation of clinically interpretable hypotheses in neurodegenerative disease research.