# Characterizing structural MR brain changes of child and adolescent bipolar disorder patients using Random Forests classification

**P-H. Yeh[1], H. Hongtu Zhu[2], M. Nicoletti[2], H. Baloch[2], J. Hatch[3], G. Zunta-Soares[2], and J. C. Soares[2]**

[1]Psychiatry, University of North Carolina, Chapel Hill, Chapel Hill, NC, United States, [2]University of North Carolina, Chapel Hill, [3]Department of Psychiatry, The University of Texas Health Science Center at San Antonio, U.S.A

**Introduction** To investigate important structural brain characteristics in classifying child and adolescent bipolar disorder (BD) patients, we used Random Forests, an improved Classification and Regression Trees (CART) machine learning method, for doing disease classifications in pathway-based analysis.

**Methods** Participants included 59 healthy controls (HC) (M/F=32/27, age 17.71 ± 2.74 y) and 59 subjects with bipolar disorder (BD) (M/F=31/28, age 17.69 ± 2.81 y). 3D-SPGR MRI brain scans were acquired at 1.5 T (TR/TE=24/5ms, 1x1x1mm$^3$, FOV= 256x256 mm$^2$, matrix=256x256mm). Cortical surface reconstructions and parcellation of brain volumes were generated with the Freesurfer software package (http://surfer.nmr.mgh.harvard.edu). The regional brain volumes and cortical thicknesses were first regressed individually with covariates, e.g. age, gender and intracranial volume. For the Random Forests *classification,* the important pathways were identified through the out-of-bag (OOB) error estimation for each fitted tree, i.e. about one-third of the original data are left out of the sample in tree construction and then used for estimating the prediction errors. The normalized proximity measures, the total count of both training and OOB divided by the numbers of trees, were used to evaluate the individual behaviors in the sample. Outlier measure, the inverse of the sum of squares of proximity, was used for the detection of pathway-based outliers. To quantify which MR structural measures are most informative for giving the correct pathway-based classification, the mean decrease accuracy (MDA), was calculated by permuting the values of each variable in the OOB cases and recording the prediction accuracy.

**Results** Fig. 1 shows the proximity matrix visualized on a multi-dimensional scaling plot, a picture of similarities among brain structural measures (regional brain volume at left, cortical thickness at right) in individuals and their respective classes (HC in blue *vs* BD in red).

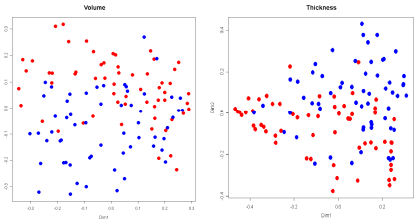


**Fig. 1**

The average OOB estimates of error rates were 46.61% (HC: 44.07%, BD: 49.15%) in brain volumes bootstrapping, and 38.98% in cortical thickness for both of HC and BD. This indicates that cortical thickness is more accurate than gray mater volume in classifying BD vs HC .

Fig. 2 shows the variable important plots for MDA (brain volume at left, cortical thickness at right), which provides a way to visualize the marginal effect of MR structural measures on class probability in Random Forests classification (HC vs BD). The important top ranked (MDA > 0.32) were the volumes of right corpus callosum, right rostral middle frontal lobe, right caudate nucleus, and left inferior parietal lobe, the cortical thickness of right isthmus cingulate gyrus, left pars triangularis, left supramarginal lobe and left lateral orbitofrontal lobe.

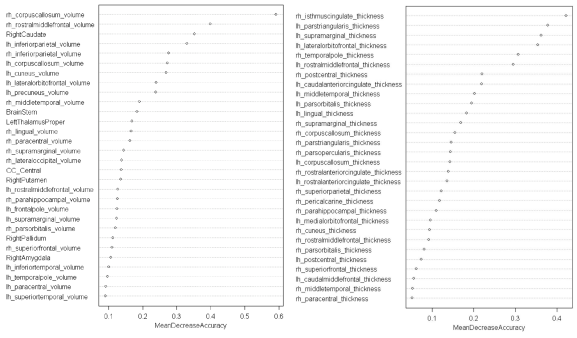


**Fig. 2**

**Discussion** Pathway-based Random Forests classification would allow us to find relevant components among different neural pathways and investigate how brain structural pathways are affected within neurocircuitry in neuropsychiatric disorders. The study provides evidence of brain structural abnormalities within the cortico-striatal-thalamic-cortical and limbic-cortico-striatal-thalamic-cortical circuits in BD.

**Acknowledgments:** This research was partly supported by MH 69774, MH 068662, RR 20571, UTHSCSA's GCRC (M01-RR-01346), and the Krus Endowed Chair in Psychiatry (UTHSCSA).