BIOS 6623: Project 3

Fall 2024

Written report and in-class presentation due date:

Monday, November 18

The formation and growth of multiple cysts in the kidneys is a characteristic of autosomal dominant polycystic kidney disease (ADPKD). Baseline height-corrected total kidney volume, which is measured using magnetic resonance imaging (MRI), has become an imaging biomarker for assessing disease progression in adult patients. However, predicting outcomes in children is more challenging. Although cysts are present in the kidneys of young ADPKD patients, kidney function may remain normal for many years. Some recent adult study demonstrated that MRI image features improved the prognostic accuracy compared with the total kidney volume. The aim of this study is to investigate whether MRI image features extracted from baseline kidney images could enhance the prediction of disease progression in young ADPKD patients. This will be achieved by using predictive models.

This project has recruited 71 young ADPKD patients. Each subject has MRI data available at baseline and after 3 years. The height-corrected total kidney volumes at baseline and after 3 years have been measured by a physician. The percentage changes of total kidney volume after 3 years compared with the baseline year has also been calculated. The subjects are labeled as fast or slow progression based on the changes of total kidney volume. Different types of image features have been extracted from the baseline MRI images.

The first task: Please develop and compare the following predictive models that use subject characteristics measured at baseline to predict the percentage changes of total kidney volume after 3 years. You may consider applying feature engineering to the predictors.

(1) A predictive model that uses the baseline height-corrected total kidney volume.
(2) A predictive model that uses image features extracted from the baseline MRI data.
(3) A predictive model that uses both the baseline height-corrected total kidney volume and image features extracted from the baseline MRI data.

The second task: Please develop and compare the following predictive models that predict whether a subject is fast or slow progression. You may consider applying feature engineering to the predictors.

(1) A predictive model that uses the baseline height-corrected total kidney volume.
(2) A predictive model that uses image features extracted from the baseline MRI data.
(3) A predictive model that uses both the baseline height-corrected total kidney volume and image features extracted from the baseline MRI data.

The coding information in the **Project3_data.csv** file is listed as follows. Note that all the image features are extracted from the baseline MRI data.

geom1/geom2: The first/second image feature based on kidney geometry information.

gabor1/gabor2/gabor3/gabor4/gabor5: The first/second/third/fourth/fifth image feature based on Gabor transform.

glcm1/glcm2: The first/second image feature based on gray level co-occurrence matrix.

txti1/txti2/txti3/txti4/txti5: The first/second/third/fourth/fifth image feature based on image textures.

lbp1/lbp2/lbp3/lbp4/lbp5: The first/second/third/fourth/fifth image feature based on local binary pattern.

tkvht_base: The baseline height-corrected total kidney volume.

tkvht_visit2: The height-corrected total kidney volume after 3 years.

progression: 1 for fast progression and 0 for slow progression.

tkvht_change: The percentage changes of total kidney volume after 3 years compared with the baseline year.