Automatic Detection and Segmentation of Brain Tumor Using Random Forest Approach

Abstract by Leonard Brenk

October 28, 2020

The main goal of this study is to develop a procedure for the detection and segmentation of brain tumors based on random forests used on multimodal MRI images. Since the early detection can enhance the medical treatment immensely the primary focus lies on the detection. In this experiment, 12 records were used each containing four volumes showing the same brain in a different scan. The pictures were preprocessed and carry connotations made by human experts. To address one specific pixel, handcrafted features were created and extended with neighborhood information. Also, the image underwent further processing steps, like normalizations and the computation of missing values. The algorithm is trained to separate between tumor, edema and negative pixels. The accuracy of it will be characterized by the Dice Score from 0 to 1.

The ensemble method used is random forest which consists of many Binary Decision Trees (BDT). One BDT describes a hierarchy of two-way decisions and can be employed to learn the classification of data into labels. When training a BDT you start with inserting a sample of input data into the root node which needs to decide for one feature whether a certain vector will be navigated to the right or left child node. For picking the feature it is been decided by, the information gain and entropy is used. The feature which will yield the highest information gain for the decision when split, will be considered at the specific node. After the tree has been trained it is being tested.

BDT tend to learn and adapt to every small detail of the training data. This is called overfitting since errors will occur when testing an unknown image. The problem of overfitting is faced with the random forest ensemble method which trains many trees and decides through a majority vote about the final label. Bagging selects a subset of data while training each tree randomly. In a random forest, a tree is then trained with only a smaller amount of input data and a randomly chosen subset of features. After training the tree the resulting volume is post-processed, where all tumor or edema pixels will be checked a second time to either validate or change the result.

The experiment concludes that using random forest for detecting brain tumors can be envisioned as a promising procedure to put to clinical use soon. The more specific results show that the sample size while training a tree can strongly influence the accuracy. Also, the Dice Score of training a tree with data set A and testing with set B might be high, while training with B and testing with A can be low. Crucial to state is that post-processing can have a serious impact on precision and accuracy. Regarding the size of a tumor, post-processing is most effective when used on a tumor in an early stage which again stresses the importance of detecting tumors early.

483 words