Source code description for CS2750 ML project

Qingyang Li

April 27, 2022

1 Packages

The project is build on jupyter notebook, program language is Python. The packages mainly used in this project as following:

- numpy (data array store and load)

- glob (for load images)

- seaborn (heatmap)

- matplotlib (graph)

- opencv-python (read images)

- tensorflow.keras (AlexNet)

- pickle (save and load models)

- slearn (ML models and data set split)

2 Run

Image set I used in this project is selected by *img_selection.ipynb*, this file will automatic select 7160 tumor negative images and 2840 tumor positive images, and 7160 negative and 7160 positive images as over-sampled images set.

2.1 data preprocess

data_preprocess.ipynb,will read images set, generate data array and store them in npy format for training and testing purpose.

2.2 Training models

For model training, run the folloing code

- Run *LR.ipynb* can train the logistic regression model

- Run *KD.ipynb* can train the KNN model

- Run *SVM.ipynb* can train the SVM model

- Run *RandomForest.ipynb* can train the Random Forest model

- Run *AlexNet.ipynb* can train the AlexNet model

AlexNet models is build by tenserflow.keras, other models are from sklearn

2.3   Result and Evaluation

Run *performance.ipynb*, execute the cells to test performance for different models, for models build on sklearn, the result display confusion matrix, precision,recall, misclassification error,SPEC,NPV, F1-score, AUROC and ROC graph for training set and testing set. For AlexNet, the result only display accuracy,recall,AUROC,number of false negatives and F1-scores