Brain MRI goodness check

Project for CSCI E-82, Fall 2021, Harvard University

By Tashrif Billah

Abstract

Brain MRI quality check is an active area of research in the medical imaging community. An MRI can be affected by patient's motion or magnetic artifacts. Research assistants have to spend many hours to visually identify whether an MRI is usable. This project aims to automate that manual labor. A 3D kernel is traversed along the 3D MRI to learn block features. Those features are represented by a continuous distribution named kernel density estimation. The distributions are compared against a pool of known good and bad images using multiple similarity metrices. Pearson correlation metric gives the best results for DIAGNOSE CTE data used in this project. An MRI is declared as a good or bad one based on its highest similarity against the particular pool of images. Bootstrapping is used to increase the number of unbalanced samples. Finally, confusion matrix is used to demonstrate results.

In [22]:
```
from IPython.display import display, Image
```

Problem

Patients can hardly remain still inside an MRI machine. Their motion affects the quality of an MRI scan. Moreover, magnetic field in the machine can induce artifacts in the MRI. These limitations sometimes make an MRI unusable that must be omitted from population analysis. They are:

- Motion artifact
- Ringing artifact
- Signal drop artifact

In [28]:
```
display(Image(filename="artifacts/motion.png", height=400, width=400))
```

- Motion artifact

  Distorted brain

In [29]:
```
display(Image(filename="artifacts/ringing.png", height=400, width=400))
```

- Ringing aritifact

  Circular line over the brain

In [30]:
```
display(Image(filename="artifacts/signal_drop.png", height=400, width=400))
```

- Signal drop artifact

  Part of back-upper brain is missing

A human has to go through several slices of a 3D MRI scan to determine the existence of artifacts. This project aims to automatically rate an MRI on a scale of 1-4, the higher the better quality. I intend to use 3-D block features and a histogram matching technique to identify bad scans by comparing them against ground truths.

Data

The MRI data used in this project are part of DIAGNOSE CTE initiative. It is a 7-Year multi-site research project to develop methods of diagnosing Chronic Traumatic Encephalopathy (CTE) during life and to examine potential risk factors for this degenerative brain disease. After MRI scans are acquired, they come to our laboratory at Brigham and Women's Hospital and Harvard Medical School. A group of multi-disciplinary scientists are tasked with analyzing them using various neuroimaging tools.

The first step before any analysis is to quality check the MRIs. A pool of postbaccalaureate research assistants observe the MRIs visually and give them a rating of [1,2,3,4]. MRIs with rating 2 or below are discarded from further study. MRI acquisition sites are advised to recall those patients for rescans. Each MRI is a 256 x 256 x 176 numeric array of intensities. Each slice is 256 x 256. Therefore there are 176 slices. The research assistants have to view them one by one to determine usability. This step is burdensome.

However, I have assembled 115 T1w and T2w MRI scans for this project. Most of them are of good quality (scores 3,4). Some of them are of bad quality (scores 1,2). They are found in data folder:

```
[tb571@pnl-z840-2 data]$ tree
.
├── 1001
│   └── raw
│       ├── 1001-t1w-xc.nrrd
│       └── 1001-t2w-xc.nrrd
├── 1002
│   └── raw
│       ├── 1002-t1w-xc.nrrd
│       └── 1002-t2w-xc.nrrd
├── 1003
│   └── raw
│       ├── 1003-t1w-xc.nrrd
│       └── 1003-t2w-xc.nrrd
...
...
```

Total data size is 2.4 GB.

Algorithm

This section is divided into three subsections:

- Preprocessing
- Featurization
- Comparision
- Decision

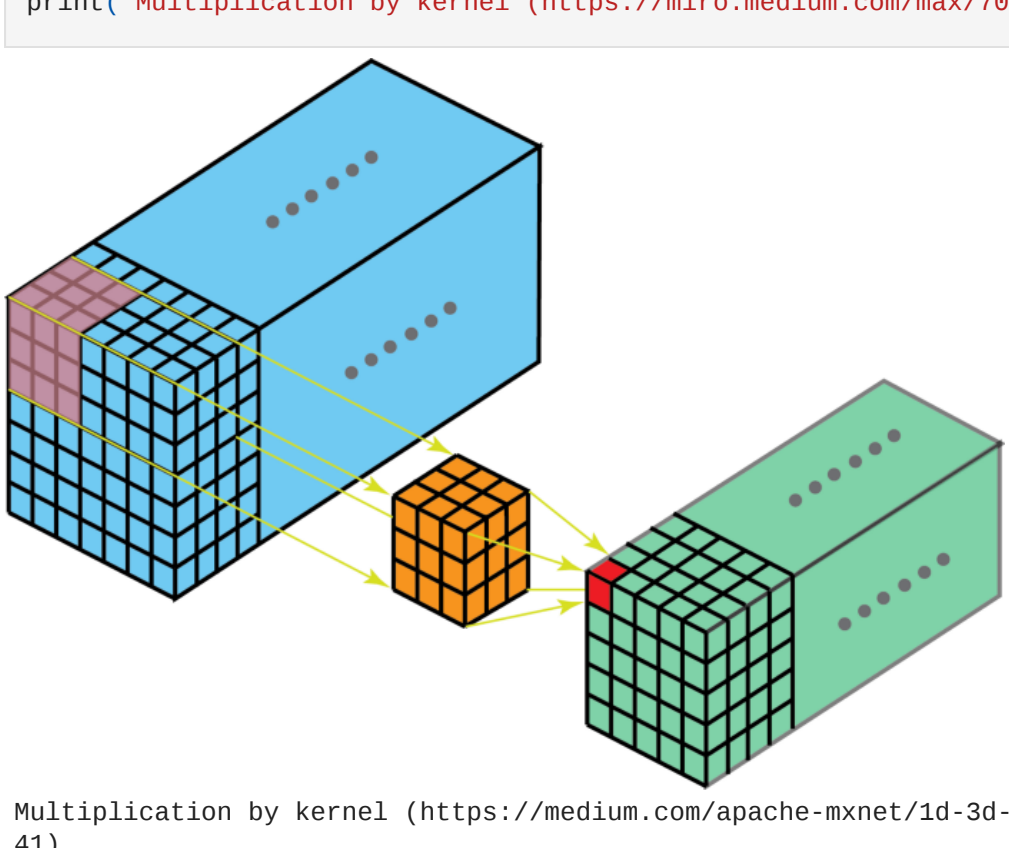
Preprocessing

All MRIs are non-linearly registered to a reference MRI. This step omits the effect of size/shape difference among brain scans. It also reduces motion artifact in the brain scans.

Featurization

- Each MRI is a 3D volume of dimension `X x Y x Z`
- Echoing the idea of convolutional neural network, a 3D kernel of dimension `nx x ny x nz` with strides `sx x sy x sz` is applied on the 3D volume
- Each multiplication of kernel and volume extracts image intensities (features) over the 3D block. There are `nx x ny x nz` data points in that block.

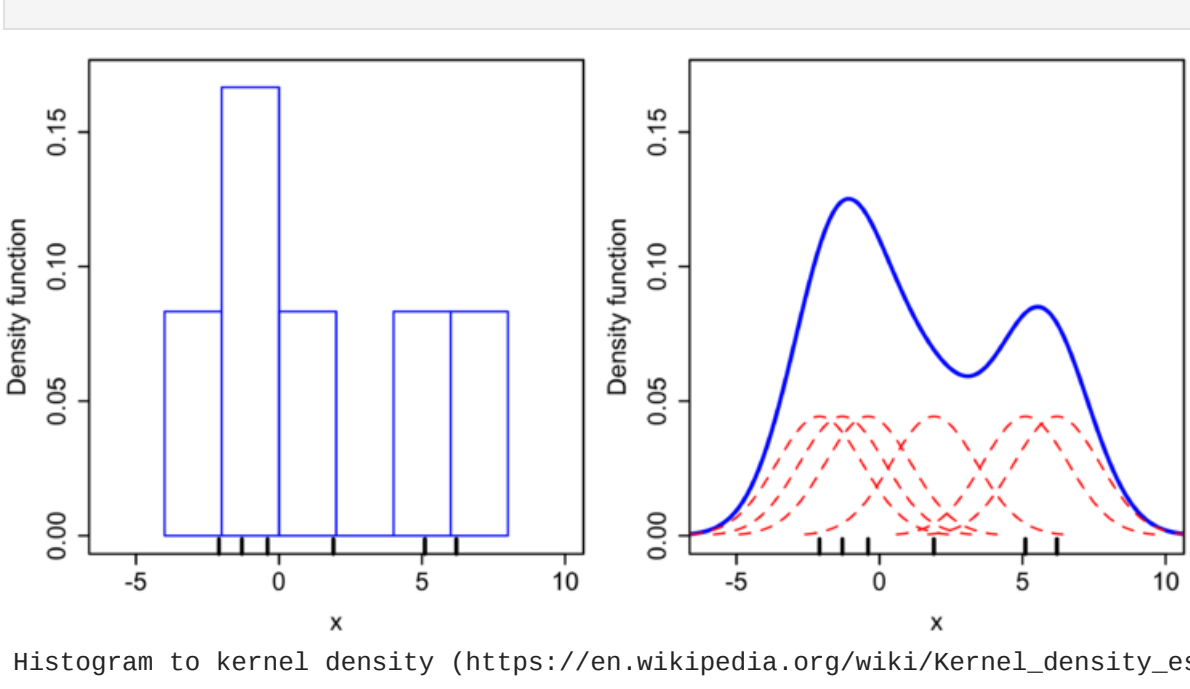
In [31]:
```
display(Image(filename="kernel.png", height=500, width=500))
print("Multiplication by kernel (https://miro.medium.com/max/700/1*wUVVgZnzBwYKgQyTBK_5sg.png)")
```



```
Multiplication by kernel (https://medium.com/apache-mxnet/1d-3d-convolutions-explained-with-ms-excel-5f88c0f359
41)
```

- A histogram is calculated for each 3D block. To facilitate further calculation, a smoothing technique e.g. kernel density estimation is applied on the 3D block. Thus, we derive a continuous distribution of features at each 3D block.

In [26]:
```
display(Image(filename="KDE.png", height=400, width=600))
print("Histogram to kernel density (https://en.wikipedia.org/wiki/Kernel_density_estimation)")
```




```
Histogram to kernel density (https://en.wikipedia.org/wiki/Kernel_density_estimation)
```

Comparison

- To decide whether an MRI is good or bad, we compare it against a pool of known good and bad ones. The latter ground truth is developed by human raters. Human raters are trained research assistants who visually observe each MRI and gives it a rating [1,2,3,4] where 1 is the most artifact affected and 4 is the perfect one.
- Image similarity is compared at that 3D block level. The algorithm is equipped to use one of the four techniques for comparing two distributions:
  - Pearson correlation coefficient
  - Bhattacharyya distance
  - Kullback–Leibler divergence
  - Mean squared error

  The higher the Pearson correlation coefficient, the more is the similarity. Meanwhile, the lower the other measures, the more is the similarity.
- Each of the techniques yields one number for each comparison. Thereby, we represent each 3D block by a single number. In total, there are X x Y x Z such numbers. They are averaged to reduce to one single number for each MRI. Thus, we have a score for each MRI when compared against a ground truth image.
- Since we have four discrete human ratings [1,2,3,4], we have four pool of ground truth images. An MRI gets a pool of comparison scores for each rating. We average them to get a single score for each rating.
- An input MRI is assigned a rating corresponding to the highest/lowest similarity score. Let's take a look at the following example:

Ground truth rating	Pool of scores	Average score	Final rating
1	[0.2, 0.2, 0.1, 0.3]	0.2	
2	[0.8, 0.8, 0.9, 0.7]	0.8	2
3	[0.1, 0.1, 0.1, 0.1]	0.1	
4	[0.9, 0.1, 0.1, 0.9]	0.5	

The MRI got rating 2 since its average similarity score (Pearson correlation coefficient) is the highest for pool of ground truth images with rating 2.

Decision

Did the above MRI pass or fail the goodness check algorithm? To answer that question, the user needs to define a `decisionFactor` in the `code/config.ini` file. The decision criterion is:

`'pass' if similarity_score > score_range / decisionFactor else 'fail'`

In the above example, `score_range=4`. If `decisionFactor=2`, then the threshold is `4/2=2`. Since it got a rating of 2, it failed the goodness check algorithm. So either the patient should be called in for a rescan or this image should be omitted from further population analysis.

Program

In this section, we shall describe how to run this project at a remote user's end. This section is divided into following subsections:

- Dependencies
- Installation
- Execution
  - Registration
  - Training and Testing

Dependencies

The dependencies are noted in `environment.yml` file:

- ANTs==2.3.0
- numpy
- nibabel
- pandas
- xlrd
- scipy
- psutil
- plumbum
- ants
- conda-build
- sklearn
- openpyxl
- pynrrd

The only outstanding dependency is ANTs. It is the state of the art tool for performing registration among neuroimages.

Installation

Please run the following commands in your terminal:

```
# install Python 3
wget https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
sh Miniconda3-latest-Linux-x86_64.sh -b -p mininconda3
source miniconda3/bin/activate

# clone this project
git clone https://github.com/tashrifbillah/csci_e82_project.git
cd csci_e82_project

# build Python environment
conda env create -f environment.yml
conda activate structQC

# set up bash environment
export ANTSPATH=$(dirname `which antsApplyTransforms`)
```

Now launch `jupyter-notebook` from that terminal to run `csci_e82_project.ipynb` at your end.

Finally, you should download the data used in this project as:

```
cd csci_e82_project/
wget https://www.dropbox.com/s/yim3blsy17krp49/csci_e82_project_data.tar.gz
tar -xzvf csci_e82_project_data.tar.gz
```

Please make sure the data got extracted to `csci_e82_project/data`:

▶ **Click here**

Execution

After installation is complete, the `code/config.ini` file needs to be updated with user's paths:

```
train_visual_qc = /home/tb571/ml_class/csci_e82_project/data/train_visual_qc.xlsx
test_visual_qc = /home/tb571/ml_class/csci_e82_project/data/test_visual_qc.xlsx
fixedimaget1 = /home/tb571/ml_class/csci_e82_project/data/1001-t1w-xc.nrrd
fixedimaget2 = /home/tb571/ml_class/csci_e82_project/data/1001-t2w-xc.nrrd
fixedmaskt1 =
fixedmaskt2 =
t1histogram = /home/tb571/ml_class/csci_e82_project/data/patch_histograms_t1.npy
t2histogram = /home/tb571/ml_class/csci_e82_project/data/patch_histograms_t2.npy
```

Now you can proceed to executing the codes.

Registration

This step is optional for grading purposes. It is the most time consuming step in the whole pipeline. Here, all MRIs are registered to a reference MRI. T1w and T2w images have separate references. We have already shipped the registered images with the submission. Performing registration again at remote user's end again will not improve results a bit. Meanwhile, logs generated from ANTs are pretty big and hardly fits to the scope of this notebook. Hence, the following registration steps have been directly executed in a Linux terminal.

```
time code/structuralQCbatch.py \
-i data/t1_list.txt \
-c data/caselist.txt \
--fixedImg data/1001-t1w-xc.nrrd \
-e data/visual_qc.xlsx \
-o data/processed \
--type t1 --register --feature --train --numThreads 16

time code/structuralQCbatch.py \
-i data/t2_list.txt \
-c data/caselist.txt \
--fixedImg data/1001-t2w-xc.nrrd \
-e data/visual_qc.xlsx \
-o data/processed \
--type t2 --register --feature --train --numThreads 16
```

It may take ~4 hours to complete registration for each modality (T1w or T2w) for the 115 subjects when four threads are used.

After the completion of registration, each MRI has a corresponding `*-reg.nii.gz` file:

```
[tb571@pnl-z840-2 data]$ tree
.
├── 1001
│   └── raw
│       ├── 1001-t1w-xc.nrrd
│       ├── 1001-t1w-xc-reg.nii.gz
│       ├── 1001-t2w-xc.nrrd
│       └── 1001-t2w-xc-reg.nii.gz
├── 1002
│   └── raw
│       ├── 1002-t1w-xc.nrrd
│       ├── 1002-t1w-xc-reg.nii.gz
│       ├── 1002-t2w-xc.nrrd
│       └── 1002-t2w-xc-reg.nii.gz
├── 1003
│   └── raw
│       ├── 1003-t1w-xc.nrrd
│       ├── 1003-t1w-xc-reg.nii.gz
│       ├── 1003-t2w-xc.nrrd
│       └── 1003-t2w-xc-reg.nii.gz
...
...
```

- Before we begin training and testing, let's take a look at the rating distribution:

In [84]:
```
import pandas as pd
from sklearn.model_selection import train_test_split

# load lists
df= pd.read_excel('data/visual_qc.xlsx')

for score in range(1,5):
    print('Score', score)
    print('T1w', df.groupby('t1 score').get_group(score).shape[0])
    print('T2w', df.groupby('t2 score').get_group(score).shape[0])

    print('')
```

```
Score 1
T1w 11
T2w 4

Score 2
T1w 9
T2w 20

Score 3
T1w 66
T2w 30

Score 4
T1w 29
T2w 61
```

- There is imbalance in rating distribution. The number of scores 1 and 2 (bad images) are quite less than that of 3 and 4 (good images). So we shall apply bootstrapping to increase the bad samples.

In [89]:
```
def partition(modality):

    # load lists
    df= pd.read_excel('visual_qc.xlsx')

    with open('caselist.txt') as f:
        cases= f.read().strip().split()

    with open(f'{modality}_reg_list.txt') as f:
        files= f.read().strip().split()


    # === Bootstrapping begins === #

    # bootstrap 1 and 2 ratings to obtain twice as many as given samples
    df1= df[df[f'{modality} score']==1]
    df1=df1.append(df1.sample(frac=1, replace=True))

    df2= df[df[f'{modality} score']==2]
    df2=df2.append(df1.sample(frac=1, replace=True))

    # update df
    df=df.append(df1)
    df=df.append(df2)

    # update cases
    cases= df['Subject ID'].values

    # update files
    template= files[0]
    for new_case in df1['Subject ID'].values:
        files.append(template.replace(str(cases[0]),str(new_case)))

    for new_case in df2['Subject ID'].values:
        files.append(template.replace(str(cases[0]),str(new_case)))

    # === Bootstrapping ends === #


    # define train test split
    train_visual, test_visual, train_cases, test_cases, train_files, test_files= \
        train_test_split(df, cases, files, test_size=0.25, stratify=df[f'{modality} score'])

    # save lists
    train_visual.to_excel(f'train_visual_qc.xlsx', index=False)
    test_visual.to_excel(f'test_visual_qc.xlsx', index=False)

    with open('train_reg_list.txt','w') as f:
        for i in train_files:
            f.write(f'{i}\n')

    with open('test_reg_list.txt','w') as f:
        for i in test_files:
            f.write(f'{i}\n')

    with open('train_cases.txt', 'w') as f:
        for i in train_cases:
            f.write(f'{i}\n')

    with open('test_cases.txt', 'w') as f:
        for i in test_cases:
            f.write(f'{i}\n')
```

Training and testing

We have already shipped the training features `t1histogram` and `t2histogram` with this submission. Performing training again will overwrite those features. If you want to use packaged training features, you can skip to the testing step. Otherwise, you can obtain new training features as follows:

T1w MRI

In [75]:
```
%cd data/
partition('t1')
%cd ..
```

```
/home/tb571/ml_class/csci_e82_project/data
/home/tb571/ml_class/csci_e82_project
```

In [77]:
```
# Train (calculate histogram)
!python code/structuralQCbatch.py \
-i data/train_reg_list.txt \
-c data/train_cases.txt \
--fixedImg data/1001-t1w-xc.nrrd \
-e data/train_visual_qc.xlsx \
-o data/processed \
--type t1 \
--feature --train \
--numThreads 16
```

```
antsApplyTransforms found
antsRegistrationSyNQuick.sh found
antsRegistration found
All executables are found, program will begin now ...
Calculating histogram of 5007-t1w-xc-reg.nii.gz
Calculating histogram of 3024-t1w-xc-reg.nii.gz
Calculating histogram of 1035-t1w-xc-reg.nii.gz
Calculating histogram of 5030-t1w-xc-reg.nii.gz
Calculating histogram of 5016-t1w-xc-reg.nii.gz
Calculating histogram of 5015-t1w-xc-reg.nii.gz
Calculating histogram of 5003-t1w-xc-reg.nii.gz
Calculating histogram of 7028-t1w-xc-reg.nii.gz
Calculating histogram of 5025-t1w-xc-reg.nii.gz
Calculating histogram of 1030-t1w-xc-reg.nii.gz
Calculating histogram of 1018-t1w-xc-reg.nii.gz
Calculating histogram of 3020-t1w-xc-reg.nii.gz
Calculating histogram of 5016-t1w-xc-reg.nii.gz
Calculating histogram of 7023-t1w-xc-reg.nii.gz
Calculating histogram of 7014-t1w-xc-reg.nii.gz
Calculating histogram of 3002-t1w-xc-reg.nii.gz
Histogram calculation successful of 1035-t1w-xc-reg.nii.gz
Calculating histogram of 7024-t1w-xc-reg.nii.gz
Histogram calculation successful of 5016-t1w-xc-reg.nii.gz
Calculating histogram of 1026-t1w-xc-reg.nii.gz
Histogram calculation successful of 7023-t1w-xc-reg.nii.gz
Calculating histogram of 7023-t1w-xc-reg.nii.gz
Histogram calculation successful of 7028-t1w-xc-reg.nii.gz
Calculating histogram of 1037-t1w-xc-reg.nii.gz
Histogram calculation successful of 5016-t1w-xc-reg.nii.gz
Calculating histogram of 3023-t1w-xc-reg.nii.gz
Histogram calculation successful of 1030-t1w-xc-reg.nii.gz
Calculating histogram of 5016-t1w-xc-reg.nii.gz
Histogram calculation successful of 5003-t1w-xc-reg.nii.gz
Calculating histogram of 5016-t1w-xc-reg.nii.gz
Histogram calculation successful of 5007-t1w-xc-reg.nii.gz
Calculating histogram of 5016-t1w-xc-reg.nii.gz
Histogram calculation successful of 5030-t1w-xc-reg.nii.gz
Calculating histogram of 7029-t1w-xc-reg.nii.gz
Histogram calculation successful of 7014-t1w-xc-reg.nii.gz
Histogram calculation successful of 5025-t1w-xc-reg.nii.gz
Calculating histogram of 3016-t1w-xc-reg.nii.gz
Calculating histogram of 1028-t1w-xc-reg.nii.gz
Histogram calculation successful of 1018-t1w-xc-reg.nii.gz
Calculating histogram of 7016-t1w-xc-reg.nii.gz
Histogram calculation successful of 5015-t1w-xc-reg.nii.gz
Calculating histogram of 3012-t1w-xc-reg.nii.gz
Histogram calculation successful of 3024-t1w-xc-reg.nii.gz
Calculating histogram of 5002-t1w-xc-reg.nii.gz
Histogram calculation successful of 3020-t1w-xc-reg.nii.gz
Calculating histogram of 1022-t1w-xc-reg.nii.gz
Histogram calculation successful of 3002-t1w-xc-reg.nii.gz
Calculating histogram of 3023-t1w-xc-reg.nii.gz
Histogram calculation successful of 7024-t1w-xc-reg.nii.gz
Calculating histogram of 1028-t1w-xc-reg.nii.gz
Histogram calculation successful of 1026-t1w-xc-reg.nii.gz
Calculating histogram of 7030-t1w-xc-reg.nii.gz
Histogram calculation successful of 1037-t1w-xc-reg.nii.gz
Calculating histogram of 7001-t1w-xc-reg.nii.gz
Histogram calculation successful of 7016-t1w-xc-reg.nii.gz
Calculating histogram of 1025-t1w-xc-reg.nii.gz
Histogram calculation successful of 7023-t1w-xc-reg.nii.gz
Calculating histogram of 5022-t1w-xc-reg.nii.gz
Histogram calculation successful of 5016-t1w-xc-reg.nii.gz
Calculating histogram of 1032-t1w-xc-reg.nii.gz
Histogram calculation successful of 7029-t1w-xc-reg.nii.gz
Calculating histogram of 5021-t1w-xc-reg.nii.gz
Histogram calculation successful of 5016-t1w-xc-reg.nii.gz
Calculating histogram of 1027-t1w-xc-reg.nii.gz
Histogram calculation successful of 1028-t1w-xc-reg.nii.gz
Calculating histogram of 5004-t1w-xc-reg.nii.gz
Histogram calculation successful of 3016-t1w-xc-reg.nii.gz
Calculating histogram of 5017-t1w-xc-reg.nii.gz
Histogram calculation successful of 5016-t1w-xc-reg.nii.gz
Calculating histogram of 1015-t1w-xc-reg.nii.gz
Histogram calculation successful of 3023-t1w-xc-reg.nii.gz
Histogram calculation successful of 1022-t1w-xc-reg.nii.gz
Calculating histogram of 1028-t1w-xc-reg.nii.gz
Calculating histogram of 7018-t1w-xc-reg.nii.gz
Histogram calculation successful of 5002-t1w-xc-reg.nii.gz
Calculating histogram of 1031-t1w-xc-reg.nii.gz
Histogram calculation successful of 3012-t1w-xc-reg.nii.gz
Calculating histogram of 5027-t1w-xc-reg.nii.gz
Histogram calculation successful of 3023-t1w-xc-reg.nii.gz
Calculating histogram of 7021-t1w-xc-reg.nii.gz
Histogram calculation successful of 1028-t1w-xc-reg.nii.gz
Calculating histogram of 5006-t1w-xc-reg.nii.gz
Histogram calculation successful of 7030-t1w-xc-reg.nii.gz
Calculating histogram of 1019-t1w-xc-reg.nii.gz
Histogram calculation successful of 5004-t1w-xc-reg.nii.gz
Calculating histogram of 3015-t1w-xc-reg.nii.gz
Histogram calculation successful of 1032-t1w-xc-reg.nii.gz
Calculating histogram of 1016-t1w-xc-reg.nii.gz
Histogram calculation successful of 7001-t1w-xc-reg.nii.gz
Calculating histogram of 1037-t1w-xc-reg.nii.gz
Histogram calculation successful of 1025-t1w-xc-reg.nii.gz
Calculating histogram of 3019-t1w-xc-reg.nii.gz
Histogram calculation successful of 5021-t1w-xc-reg.nii.gz
Calculating histogram of 5020-t1w-xc-reg.nii.gz
Histogram calculation successful of 5022-t1w-xc-reg.nii.gz
Calculating histogram of 1005-t1w-xc-reg.nii.gz
Histogram calculation successful of 1027-t1w-xc-reg.nii.gz
Calculating histogram of 5016-t1w-xc-reg.nii.gz
Histogram calculation successful of 1015-t1w-xc-reg.nii.gz
Calculating histogram of 3021-t1w-xc-reg.nii.gz
Histogram calculation successful of 7018-t1w-xc-reg.nii.gz
Calculating histogram of 1021-t1w-xc-reg.nii.gz
Histogram calculation successful of 5017-t1w-xc-reg.nii.gz
Calculating histogram of 1022-t1w-xc-reg.nii.gz
Histogram calculation successful of 1031-t1w-xc-reg.nii.gz
Calculating histogram of 7002-t1w-xc-reg.nii.gz
Histogram calculation successful of 5027-t1w-xc-reg.nii.gz
Calculating histogram of 5024-t1w-xc-reg.nii.gz
Histogram calculation successful of 1028-t1w-xc-reg.nii.gz
Calculating histogram of 7031-t1w-xc-reg.nii.gz
Histogram calculation successful of 7021-t1w-xc-reg.nii.gz
Calculating histogram of 3012-t1w-xc-reg.nii.gz
Histogram calculation successful of 5006-t1w-xc-reg.nii.gz
Calculating histogram of 7028-t1w-xc-reg.nii.gz
Histogram calculation successful of 1019-t1w-xc-reg.nii.gz
Calculating histogram of 7028-t1w-xc-reg.nii.gz
Histogram calculation successful of 1037-t1w-xc-reg.nii.gz
Calculating histogram of 7019-t1w-xc-reg.nii.gz
Histogram calculation successful of 3015-t1w-xc-reg.nii.gz
Calculating histogram of 5019-t1w-xc-reg.nii.gz
Histogram calculation successful of 1005-t1w-xc-reg.nii.gz
Calculating histogram of 1017-t1w-xc-reg.nii.gz
Histogram calculation successful of 1016-t1w-xc-reg.nii.gz
Calculating histogram of 7011-t1w-xc-reg.nii.gz
Histogram calculation successful of 5020-t1w-xc-reg.nii.gz
Calculating histogram of 5023-t1w-xc-reg.nii.gz
Histogram calculation successful of 5016-t1w-xc-reg.nii.gz
Calculating histogram of 5016-t1w-xc-reg.nii.gz
Histogram calculation successful of 1022-t1w-xc-reg.nii.gz
Calculating histogram of 7013-t1w-xc-reg.nii.gz
Histogram calculation successful of 1021-t1w-xc-reg.nii.gz
Calculating histogram of 5016-t1w-xc-reg.nii.gz
Histogram calculation successful of 3019-t1w-xc-reg.nii.gz
Calculating histogram of 5029-t1w-xc-reg.nii.gz
Histogram calculation successful of 7031-t1w-xc-reg.nii.gz
Calculating histogram of 7015-t1w-xc-reg.nii.gz
Histogram calculation successful of 7002-t1w-xc-reg.nii.gz
Calculating histogram of 7018-t1w-xc-reg.nii.gz
Histogram calculation successful of 3021-t1w-xc-reg.nii.gz
Calculating histogram of 1020-t1w-xc-reg.nii.gz
Histogram calculation successful of 5024-t1w-xc-reg.nii.gz
Calculating histogram of 5012-t1w-xc-reg.nii.gz
Histogram calculation successful of 3012-t1w-xc-reg.nii.gz
Calculating histogram of 1001-t1w-xc-reg.nii.gz
Histogram calculation successful of 7028-t1w-xc-reg.nii.gz
Calculating histogram of 1028-t1w-xc-reg.nii.gz
Histogram calculation successful of 7019-t1w-xc-reg.nii.gz
Calculating histogram of 5014-t1w-xc-reg.nii.gz
Histogram calculation successful of 7028-t1w-xc-reg.nii.gz
Calculating histogram of 7017-t1w-xc-reg.nii.gz
Histogram calculation successful of 1017-t1w-xc-reg.nii.gz
Calculating histogram of 3017-t1w-xc-reg.nii.gz
Histogram calculation successful of 5019-t1w-xc-reg.nii.gz
Calculating histogram of 1028-t1w-xc-reg.nii.gz
Histogram calculation successful of 5023-t1w-xc-reg.nii.gz
Calculating histogram of 1023-t1w-xc-reg.nii.gz
Histogram calculation successful of 7013-t1w-xc-reg.nii.gz
Calculating histogram of 5008-t1w-xc-reg.nii.gz
Histogram calculation successful of 7011-t1w-xc-reg.nii.gz
Calculating histogram of 3012-t1w-xc-reg.nii.gz
Histogram calculation successful of 5016-t1w-xc-reg.nii.gz
Calculating histogram of 1011-t1w-xc-reg.nii.gz
Histogram calculation successful of 5029-t1w-xc-reg.nii.gz
Calculating histogram of 5006-t1w-xc-reg.nii.gz
Histogram calculation successful of 5016-t1w-xc-reg.nii.gz
Calculating histogram of 5016-t1w-xc-reg.nii.gz
Histogram calculation successful of 7018-t1w-xc-reg.nii.gz
Calculating histogram of 1002-t1w-xc-reg.nii.gz
Histogram calculation successful of 5012-t1w-xc-reg.nii.gz
Calculating histogram of 1012-t1w-xc-reg.nii.gz
Histogram calculation successful of 7015-t1w-xc-reg.nii.gz
Calculating histogram of 1006-t1w-xc-reg.nii.gz
Histogram calculation successful of 1020-t1w-xc-reg.nii.gz
Calculating histogram of 3012-t1w-xc-reg.nii.gz
Histogram calculation successful of 1001-t1w-xc-reg.nii.gz
Calculating histogram of 5009-t1w-xc-reg.nii.gz
Histogram calculation successful of 7017-t1w-xc-reg.nii.gz
Histogram calculation successful of 5014-t1w-xc-reg.nii.gz
Calculating histogram of 7029-t1w-xc-reg.nii.gz
Calculating histogram of 1017-t1w-xc-reg.nii.gz
Histogram calculation successful of 1028-t1w-xc-reg.nii.gz
Calculating histogram of 7004-t1w-xc-reg.nii.gz
Histogram calculation successful of 3017-t1w-xc-reg.nii.gz
Calculating histogram of 7020-t1w-xc-reg.nii.gz
Histogram calculation successful of 1028-t1w-xc-reg.nii.gz
Calculating histogram of 1034-t1w-xc-reg.nii.gz
Histogram calculation successful of 1011-t1w-xc-reg.nii.gz
Calculating histogram of 3011-t1w-xc-reg.nii.gz
Histogram calculation successful of 1023-t1w-xc-reg.nii.gz
Calculating histogram of 7007-t1w-xc-reg.nii.gz
Histogram calculation successful of 1002-t1w-xc-reg.nii.gz
Calculating histogram of 7026-t1w-xc-reg.nii.gz
Histogram calculation successful of 5008-t1w-xc-reg.nii.gz
Calculating histogram of 5011-t1w-xc-reg.nii.gz
Histogram calculation successful of 5016-t1w-xc-reg.nii.gz
Calculating histogram of 7005-t1w-xc-reg.nii.gz
Histogram calculation successful of 5006-t1w-xc-reg.nii.gz
Calculating histogram of 1028-t1w-xc-reg.nii.gz
Histogram calculation successful of 1006-t1w-xc-reg.nii.gz
Calculating histogram of 7024-t1w-xc-reg.nii.gz
Histogram calculation successful of 3012-t1w-xc-reg.nii.gz
Calculating histogram of 5013-t1w-xc-reg.nii.gz
Histogram calculation successful of 1012-t1w-xc-reg.nii.gz
Calculating histogram of 7027-t1w-xc-reg.nii.gz
Histogram calculation successful of 3012-t1w-xc-reg.nii.gz
Calculating histogram of 1008-t1w-xc-reg.nii.gz
Histogram calculation successful of 5009-t1w-xc-reg.nii.gz
Histogram calculation successful of 1017-t1w-xc-reg.nii.gz
Calculating histogram of 3012-t1w-xc-reg.nii.gz
Histogram calculation successful of 7029-t1w-xc-reg.nii.gz
Calculating histogram of 1003-t1w-xc-reg.nii.gz
Histogram calculation successful of 7020-t1w-xc-reg.nii.gz
Calculating histogram of 1029-t1w-xc-reg.nii.gz
Histogram calculation successful of 1034-t1w-xc-reg.nii.gz
Calculating histogram of 3026-t1w-xc-reg.nii.gz
Histogram calculation successful of 7026-t1w-xc-reg.nii.gz
Calculating histogram of 5016-t1w-xc-reg.nii.gz
Histogram calculation successful of 3011-t1w-xc-reg.nii.gz
Calculating histogram of 3012-t1w-xc-reg.nii.gz
Histogram calculation successful of 7007-t1w-xc-reg.nii.gz
Calculating histogram of 3014-t1w-xc-reg.nii.gz
Histogram calculation successful of 7004-t1w-xc-reg.nii.gz
Calculating histogram of 3003-t1w-xc-reg.nii.gz
Histogram calculation successful of 5013-t1w-xc-reg.nii.gz
Calculating histogram of 7008-t1w-xc-reg.nii.gz
Histogram calculation successful of 7024-t1w-xc-reg.nii.gz
Calculating histogram of 3003-t1w-xc-reg.nii.gz
Histogram calculation successful of 5011-t1w-xc-reg.nii.gz
Calculating histogram of 1028-t1w-xc-reg.nii.gz
Histogram calculation successful of 1028-t1w-xc-reg.nii.gz
Calculating histogram of 3025-t1w-xc-reg.nii.gz
Histogram calculation successful of 7027-t1w-xc-reg.nii.gz
Calculating histogram of 3002-t1w-xc-reg.nii.gz
Histogram calculation successful of 7005-t1w-xc-reg.nii.gz
Calculating histogram of 7023-t1w-xc-reg.nii.gz
Histogram calculation successful of 1008-t1w-xc-reg.nii.gz
Calculating histogram of 3002-t1w-xc-reg.nii.gz
Histogram calculation successful of 1029-t1w-xc-reg.nii.gz
Histogram calculation successful of 1003-t1w-xc-reg.nii.gz
Histogram calculation successful of 3026-t1w-xc-reg.nii.gz
Histogram calculation successful of 3012-t1w-xc-reg.nii.gz
Histogram calculation successful of 5016-t1w-xc-reg.nii.gz
Histogram calculation successful of 3012-t1w-xc-reg.nii.gz
Histogram calculation successful of 3003-t1w-xc-reg.nii.gz
Histogram calculation successful of 3014-t1w-xc-reg.nii.gz
Histogram calculation successful of 7008-t1w-xc-reg.nii.gz
Histogram calculation successful of 1028-t1w-xc-reg.nii.gz
Histogram calculation successful of 3003-t1w-xc-reg.nii.gz
Histogram calculation successful of 7023-t1w-xc-reg.nii.gz
Histogram calculation successful of 3025-t1w-xc-reg.nii.gz
Histogram calculation successful of 3002-t1w-xc-reg.nii.gz
Histogram calculation successful of 3002-t1w-xc-reg.nii.gz
Completed histogram calculation of all the subjects.
Time taken in histogram calculation 99.71376132965088 seconds
Time taken in feature extraction for training 99.87679529190063 seconds
```

In [78]:
```
# Test (predict goodness)
!python code/structuralQCbatch.py \
-i data/test_reg_list.txt \
-c data/test_cases.txt \
--fixedImg data/1001-t1w-xc.nrrd \
-e data/test_visual_qc.xlsx \
-o data/processed \
--type t1 \
--numThreads 16
```

```
antsApplyTransforms found
antsRegistrationSyNQuick.sh found
antsRegistration found
All executables are found, program will begin now ...
/home/tb571/ml_class/csci_e82_project/data/1028/raw/1028-t1w-xc-reg.nii.gz
/home/tb571/ml_class/csci_e82_project/data/7028/raw/7028-t1w-xc-reg.nii.gz
Registered image found ...
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/7023/raw/7023-t1w-xc-reg.nii.gz
/home/tb571/ml_class/csci_e82_project/data/5021/raw/5021-t1w-xc-reg.nii.gz
Registered image found ...
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/1036/raw/1036-t1w-xc-reg.nii.gz
/home/tb571/ml_class/csci_e82_project/data/3005/raw/3005-t1w-xc-reg.nii.gz
Registered image found ...
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/1004/raw/1004-t1w-xc-reg.nii.gz
/home/tb571/ml_class/csci_e82_project/data/7012/raw/7012-t1w-xc-reg.nii.gz
Registered image found ...
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/7028/raw/7028-t1w-xc-reg.nii.gz
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/5010/raw/5010-t1w-xc-reg.nii.gz
/home/tb571/ml_class/csci_e82_project/data/1007/raw/1007-t1w-xc-reg.nii.gz
Registered image found ...
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/7006/raw/7006-t1w-xc-reg.nii.gz
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/7027/raw/7027-t1w-xc-reg.nii.gz
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/3027/raw/3027-t1w-xc-reg.nii.gz
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/5005/raw/5005-t1w-xc-reg.nii.gz
/home/tb571/ml_class/csci_e82_project/data/7022/raw/7022-t1w-xc-reg.nii.gz
Registered image found ...
Registered image found ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/3018/raw/3018-t1w-xc-reg.nii.gz
Registered image found ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/7010/raw/7010-t1w-xc-reg.nii.gz
Registered image found ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/5016/raw/5016-t1w-xc-reg.nii.gz
Registered image found ...
1 fail
/home/tb571/ml_class/csci_e82_project/data/3001/raw/3001-t1w-xc-reg.nii.gz
Registered image found ...
1 fail
/home/tb571/ml_class/csci_e82_project/data/3023/raw/3023-t1w-xc-reg.nii.gz
Registered image found ...
1 fail
/home/tb571/ml_class/csci_e82_project/data/7027/raw/7027-t1w-xc-reg.nii.gz
Registered image found ...
3 pass
/home/tb571/ml_class/csci_e82_project/data/5028/raw/5028-t1w-xc-reg.nii.gz
Registered image found ...
3 pass
/home/tb571/ml_class/csci_e82_project/data/3006/raw/3006-t1w-xc-reg.nii.gz
Registered image found ...
1 fail
/home/tb571/ml_class/csci_e82_project/data/7002/raw/7002-t1w-xc-reg.nii.gz
Registered image found ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/5006/raw/5006-t1w-xc-reg.nii.gz
Registered image found ...
1 fail
/home/tb571/ml_class/csci_e82_project/data/1031/raw/1031-t1w-xc-reg.nii.gz
Registered image found ...
2 fail
/home/tb571/ml_class/csci_e82_project/data/3010/raw/3010-t1w-xc-reg.nii.gz
Registered image found ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/5028/raw/5028-t1w-xc-reg.nii.gz
Registered image found ...
Checking quality ...
Checking quality ...
Checking quality ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/3013/raw/3013-t1w-xc-reg.nii.gz
Registered image found ...
Checking quality ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/7032/raw/7032-t1w-xc-reg.nii.gz
Registered image found ...
Checking quality ...
Checking quality ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/5016/raw/5016-t1w-xc-reg.nii.gz
Registered image found ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/5001/raw/5001-t1w-xc-reg.nii.gz
Registered image found ...
1 fail
/home/tb571/ml_class/csci_e82_project/data/5006/raw/5006-t1w-xc-reg.nii.gz
Registered image found ...
3 pass
/home/tb571/ml_class/csci_e82_project/data/3008/raw/3008-t1w-xc-reg.nii.gz
Registered image found ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/1024/raw/1024-t1w-xc-reg.nii.gz
Registered image found ...
1 fail
/home/tb571/ml_class/csci_e82_project/data/7003/raw/7003-t1w-xc-reg.nii.gz
Registered image found ...
3 pass
/home/tb571/ml_class/csci_e82_project/data/7028/raw/7028-t1w-xc-reg.nii.gz
Registered image found ...
1 fail
/home/tb571/ml_class/csci_e82_project/data/7028/raw/7028-t1w-xc-reg.nii.gz
Registered image found ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
2 fail
/home/tb571/ml_class/csci_e82_project/data/3023/raw/3023-t1w-xc-reg.nii.gz
Registered image found ...
2 fail
/home/tb571/ml_class/csci_e82_project/data/3009/raw/3009-t1w-xc-reg.nii.gz
Registered image found ...
1 fail
/home/tb571/ml_class/csci_e82_project/data/1010/raw/1010-t1w-xc-reg.nii.gz
Registered image found ...
Checking quality ...
Checking quality ...
Checking quality ...
3 pass
4 pass
3 pass
1 fail
4 pass
3 pass
Checking quality ...
Checking quality ...
Checking quality ...
4 pass
3 pass
1 fail
1 fail
4 pass
3 pass
4 pass
3 pass
1 fail
1 fail
Time taken in quality checking 486.03051495552063 seconds
```

In [79]:

```
%cat data/processed/confusion_matrix.txt
```

```
Confusion matrix (columns are true labels, rows are predicted labels):
Visual discrete scores are:  [1, 2, 3, 4]
[[14  0  0  0]
 [ 0  3  2  0]
 [ 0  0  8  8]
 [ 0  0  0  7]]
```

Observation

Presence of some off-diagonal elements in the confusion matrix tells us that not all human ratings have been accurately recovered. But we care more about the pass/fail decision where decision threshold of 2 comes into play. This threshold allows us to consolidate each quadrant of the above matrix as:

```
[[17  2]
 [0  23]]
```

From the consolidated confusion matrix, we conclude that the algorithm has mostly been able to decide which MRIs are good and which are bad ones. The number of misclassification is only 2.

Moreover, a `-t*w-xc-reg-quality.txt` file corresponding to each MRI has been created with the decision to accept/reject:

```
Predicted score: 3 (1 being worst, 4 being best)
Decision: pass
```

Furthermore, a combined scores files has been created:

```
[tb571@pnl-z840-2 data]$ cat processed/t1_QC_scores.csv
Case #,"Predicted score (1 being worst, 4 being best)",Quality
7024,4,pass
7027,1,fail
3021,1,fail
3005,4,pass
...
...
```

Any of the above can be considered to realize algorithm performance.

T2w MRI

In [93]:

```
%cd data/
partition('t2')
%cd ..
```

```
/home/tb571/ml_class/csci_e82_project/data
/home/tb571/ml_class/csci_e82_project
```

In [97]:

```
# Train (calculate histogram)
!python code/structuralQCbatch.py \
-i data/train_reg_list.txt \
-c data/train_cases.txt \
--fixedImg data/1001-t2w-xc.nrrd \
-e data/train_visual_qc.xlsx \
-o data/processed \
--type t2 \
--feature --train \
--numThreads 16
```

```
antsApplyTransforms found
antsRegistrationSyNQuick.sh found
antsRegistration found
All executables are found, program will begin now ...
Calculating histogram of 7013-t2w-xc-reg.nii.gz
Calculating histogram of 5019-t2w-xc-reg.nii.gz
Calculating histogram of 7018-t2w-xc-reg.nii.gz
Calculating histogram of 5001-t2w-xc-reg.nii.gz
Calculating histogram of 3023-t2w-xc-reg.nii.gz
Calculating histogram of 5009-t2w-xc-reg.nii.gz
Calculating histogram of 1028-t2w-xc-reg.nii.gz
Calculating histogram of 3002-t2w-xc-reg.nii.gz
Calculating histogram of 1034-t2w-xc-reg.nii.gz
Calculating histogram of 1003-t2w-xc-reg.nii.gz
Calculating histogram of 5023-t2w-xc-reg.nii.gz
Calculating histogram of 5005-t2w-xc-reg.nii.gz
Calculating histogram of 1018-t2w-xc-reg.nii.gz
Calculating histogram of 1011-t2w-xc-reg.nii.gz
Calculating histogram of 7012-t2w-xc-reg.nii.gz
Calculating histogram of 7024-t2w-xc-reg.nii.gz
Histogram calculation successful of 1011-t2w-xc-reg.nii.gz
Calculating histogram of 1006-t2w-xc-reg.nii.gz
Histogram calculation successful of 7024-t2w-xc-reg.nii.gz
Calculating histogram of 1020-t2w-xc-reg.nii.gz
Histogram calculation successful of 5023-t2w-xc-reg.nii.gz
Calculating histogram of 7011-t2w-xc-reg.nii.gz
Histogram calculation successful of 1003-t2w-xc-reg.nii.gz
Calculating histogram of 5030-t2w-xc-reg.nii.gz
Histogram calculation successful of 5005-t2w-xc-reg.nii.gz
Calculating histogram of 7008-t2w-xc-reg.nii.gz
Histogram calculation successful of 1034-t2w-xc-reg.nii.gz
Calculating histogram of 5024-t2w-xc-reg.nii.gz
Histogram calculation successful of 5019-t2w-xc-reg.nii.gz
Calculating histogram of 3021-t2w-xc-reg.nii.gz
Histogram calculation successful of 7018-t2w-xc-reg.nii.gz
Histogram calculation successful of 7012-t2w-xc-reg.nii.gz
Calculating histogram of 1004-t2w-xc-reg.nii.gz
Calculating histogram of 5028-t2w-xc-reg.nii.gz
Histogram calculation successful of 7013-t2w-xc-reg.nii.gz
Calculating histogram of 1008-t2w-xc-reg.nii.gz
Histogram calculation successful of 5009-t2w-xc-reg.nii.gz
Calculating histogram of 5021-t2w-xc-reg.nii.gz
Histogram calculation successful of 1018-t2w-xc-reg.nii.gz
Calculating histogram of 1022-t2w-xc-reg.nii.gz
Histogram calculation successful of 1028-t2w-xc-reg.nii.gz
Calculating histogram of 1022-t2w-xc-reg.nii.gz
Histogram calculation successful of 5001-t2w-xc-reg.nii.gz
Calculating histogram of 1037-t2w-xc-reg.nii.gz
Histogram calculation successful of 3023-t2w-xc-reg.nii.gz
Calculating histogram of 1028-t2w-xc-reg.nii.gz
Histogram calculation successful of 3002-t2w-xc-reg.nii.gz
Calculating histogram of 7029-t2w-xc-reg.nii.gz
Histogram calculation successful of 1006-t2w-xc-reg.nii.gz
Calculating histogram of 5011-t2w-xc-reg.nii.gz
Histogram calculation successful of 7008-t2w-xc-reg.nii.gz
Histogram calculation successful of 1020-t2w-xc-reg.nii.gz
Calculating histogram of 7004-t2w-xc-reg.nii.gz
Calculating histogram of 1025-t2w-xc-reg.nii.gz
Histogram calculation successful of 1004-t2w-xc-reg.nii.gz
Calculating histogram of 5007-t2w-xc-reg.nii.gz
Histogram calculation successful of 5030-t2w-xc-reg.nii.gz
Calculating histogram of 5016-t2w-xc-reg.nii.gz
Histogram calculation successful of 1037-t2w-xc-reg.nii.gz
Calculating histogram of 1035-t2w-xc-reg.nii.gz
Histogram calculation successful of 1022-t2w-xc-reg.nii.gz
Histogram calculation successful of 7011-t2w-xc-reg.nii.gz
Calculating histogram of 1008-t2w-xc-reg.nii.gz
Calculating histogram of 1032-t2w-xc-reg.nii.gz
Histogram calculation successful of 5024-t2w-xc-reg.nii.gz
Calculating histogram of 1024-t2w-xc-reg.nii.gz
Histogram calculation successful of 7029-t2w-xc-reg.nii.gz
Histogram calculation successful of 1022-t2w-xc-reg.nii.gz
Calculating histogram of 7010-t2w-xc-reg.nii.gz
Calculating histogram of 1001-t2w-xc-reg.nii.gz
Histogram calculation successful of 5028-t2w-xc-reg.nii.gz
Calculating histogram of 1028-t2w-xc-reg.nii.gz
Histogram calculation successful of 5021-t2w-xc-reg.nii.gz
Calculating histogram of 7002-t2w-xc-reg.nii.gz
Histogram calculation successful of 1008-t2w-xc-reg.nii.gz
Calculating histogram of 1002-t2w-xc-reg.nii.gz
Histogram calculation successful of 1028-t2w-xc-reg.nii.gz
Calculating histogram of 5004-t2w-xc-reg.nii.gz
Histogram calculation successful of 3021-t2w-xc-reg.nii.gz
Calculating histogram of 5017-t2w-xc-reg.nii.gz
Histogram calculation successful of 5011-t2w-xc-reg.nii.gz
Calculating histogram of 1028-t2w-xc-reg.nii.gz
Histogram calculation successful of 1035-t2w-xc-reg.nii.gz
Calculating histogram of 5016-t2w-xc-reg.nii.gz
Histogram calculation successful of 1002-t2w-xc-reg.nii.gz
Calculating histogram of 1029-t2w-xc-reg.nii.gz
Histogram calculation successful of 1025-t2w-xc-reg.nii.gz
Histogram calculation successful of 1032-t2w-xc-reg.nii.gz
Calculating histogram of 3008-t2w-xc-reg.nii.gz
Calculating histogram of 1015-t2w-xc-reg.nii.gz
Histogram calculation successful of 1024-t2w-xc-reg.nii.gz
Calculating histogram of 5021-t2w-xc-reg.nii.gz
Histogram calculation successful of 5007-t2w-xc-reg.nii.gz
Calculating histogram of 1019-t2w-xc-reg.nii.gz
Histogram calculation successful of 5016-t2w-xc-reg.nii.gz
Calculating histogram of 5022-t2w-xc-reg.nii.gz
Histogram calculation successful of 1008-t2w-xc-reg.nii.gz
Calculating histogram of 1012-t2w-xc-reg.nii.gz
Histogram calculation successful of 7010-t2w-xc-reg.nii.gz
Calculating histogram of 1036-t2w-xc-reg.nii.gz
Histogram calculation successful of 5004-t2w-xc-reg.nii.gz
Calculating histogram of 5029-t2w-xc-reg.nii.gz
Histogram calculation successful of 1028-t2w-xc-reg.nii.gz
Calculating histogram of 1028-t2w-xc-reg.nii.gz
Histogram calculation successful of 7002-t2w-xc-reg.nii.gz
Calculating histogram of 1016-t2w-xc-reg.nii.gz
Histogram calculation successful of 7004-t2w-xc-reg.nii.gz
Calculating histogram of 7030-t2w-xc-reg.nii.gz
Histogram calculation successful of 5017-t2w-xc-reg.nii.gz
Calculating histogram of 7032-t2w-xc-reg.nii.gz
Histogram calculation successful of 1001-t2w-xc-reg.nii.gz
Calculating histogram of 7019-t2w-xc-reg.nii.gz
Histogram calculation successful of 1028-t2w-xc-reg.nii.gz
Calculating histogram of 1028-t2w-xc-reg.nii.gz
Histogram calculation successful of 1029-t2w-xc-reg.nii.gz
Calculating histogram of 5016-t2w-xc-reg.nii.gz
Histogram calculation successful of 3008-t2w-xc-reg.nii.gz
Calculating histogram of 7005-t2w-xc-reg.nii.gz
Histogram calculation successful of 5016-t2w-xc-reg.nii.gz
Calculating histogram of 3013-t2w-xc-reg.nii.gz
Histogram calculation successful of 1015-t2w-xc-reg.nii.gz
Calculating histogram of 3016-t2w-xc-reg.nii.gz
Histogram calculation successful of 1019-t2w-xc-reg.nii.gz
Calculating histogram of 5014-t2w-xc-reg.nii.gz
Histogram calculation successful of 5021-t2w-xc-reg.nii.gz
Calculating histogram of 5016-t2w-xc-reg.nii.gz
Histogram calculation successful of 5029-t2w-xc-reg.nii.gz
Calculating histogram of 5016-t2w-xc-reg.nii.gz
Histogram calculation successful of 5022-t2w-xc-reg.nii.gz
Calculating histogram of 7016-t2w-xc-reg.nii.gz
Histogram calculation successful of 1012-t2w-xc-reg.nii.gz
Calculating histogram of 5028-t2w-xc-reg.nii.gz
Histogram calculation successful of 1036-t2w-xc-reg.nii.gz
Calculating histogram of 7029-t2w-xc-reg.nii.gz
Histogram calculation successful of 1028-t2w-xc-reg.nii.gz
Calculating histogram of 1002-t2w-xc-reg.nii.gz
Histogram calculation successful of 7032-t2w-xc-reg.nii.gz
Calculating histogram of 5016-t2w-xc-reg.nii.gz
Histogram calculation successful of 1016-t2w-xc-reg.nii.gz
Calculating histogram of 7021-t2w-xc-reg.nii.gz
Histogram calculation successful of 7019-t2w-xc-reg.nii.gz
Calculating histogram of 3014-t2w-xc-reg.nii.gz
Histogram calculation successful of 7030-t2w-xc-reg.nii.gz
Calculating histogram of 1017-t2w-xc-reg.nii.gz
Histogram calculation successful of 1028-t2w-xc-reg.nii.gz
Calculating histogram of 3015-t2w-xc-reg.nii.gz
Histogram calculation successful of 5016-t2w-xc-reg.nii.gz
Calculating histogram of 3001-t2w-xc-reg.nii.gz
Histogram calculation successful of 3016-t2w-xc-reg.nii.gz
Calculating histogram of 1028-t2w-xc-reg.nii.gz
Histogram calculation successful of 1002-t2w-xc-reg.nii.gz
Calculating histogram of 7003-t2w-xc-reg.nii.gz
Histogram calculation successful of 7005-t2w-xc-reg.nii.gz
Calculating histogram of 3027-t2w-xc-reg.nii.gz
Histogram calculation successful of 5014-t2w-xc-reg.nii.gz
Calculating histogram of 7017-t2w-xc-reg.nii.gz
Histogram calculation successful of 3013-t2w-xc-reg.nii.gz
Calculating histogram of 3012-t2w-xc-reg.nii.gz
Histogram calculation successful of 5016-t2w-xc-reg.nii.gz
Calculating histogram of 3010-t2w-xc-reg.nii.gz
Histogram calculation successful of 7016-t2w-xc-reg.nii.gz
Calculating histogram of 1032-t2w-xc-reg.nii.gz
Histogram calculation successful of 5016-t2w-xc-reg.nii.gz
Calculating histogram of 1030-t2w-xc-reg.nii.gz
Histogram calculation successful of 5028-t2w-xc-reg.nii.gz
Calculating histogram of 5008-t2w-xc-reg.nii.gz
Histogram calculation successful of 7029-t2w-xc-reg.nii.gz
Calculating histogram of 3027-t2w-xc-reg.nii.gz
Histogram calculation successful of 1017-t2w-xc-reg.nii.gz
Calculating histogram of 3020-t2w-xc-reg.nii.gz
Histogram calculation successful of 5016-t2w-xc-reg.nii.gz
Calculating histogram of 1010-t2w-xc-reg.nii.gz
Histogram calculation successful of 7021-t2w-xc-reg.nii.gz
Calculating histogram of 3005-t2w-xc-reg.nii.gz
Histogram calculation successful of 3014-t2w-xc-reg.nii.gz
Calculating histogram of 7015-t2w-xc-reg.nii.gz
Histogram calculation successful of 3015-t2w-xc-reg.nii.gz
Calculating histogram of 7027-t2w-xc-reg.nii.gz
Histogram calculation successful of 3001-t2w-xc-reg.nii.gz
Calculating histogram of 7001-t2w-xc-reg.nii.gz
Histogram calculation successful of 1028-t2w-xc-reg.nii.gz
Calculating histogram of 5002-t2w-xc-reg.nii.gz
Histogram calculation successful of 3027-t2w-xc-reg.nii.gz
Calculating histogram of 7022-t2w-xc-reg.nii.gz
Histogram calculation successful of 7017-t2w-xc-reg.nii.gz
Calculating histogram of 3009-t2w-xc-reg.nii.gz
Histogram calculation successful of 1032-t2w-xc-reg.nii.gz
Calculating histogram of 5011-t2w-xc-reg.nii.gz
Histogram calculation successful of 3010-t2w-xc-reg.nii.gz
Histogram calculation successful of 5008-t2w-xc-reg.nii.gz
Calculating histogram of 3018-t2w-xc-reg.nii.gz
Calculating histogram of 7006-t2w-xc-reg.nii.gz
Histogram calculation successful of 3012-t2w-xc-reg.nii.gz
Calculating histogram of 1023-t2w-xc-reg.nii.gz
Histogram calculation successful of 1030-t2w-xc-reg.nii.gz
Histogram calculation successful of 7003-t2w-xc-reg.nii.gz
Histogram calculation successful of 3027-t2w-xc-reg.nii.gz
Histogram calculation successful of 1010-t2w-xc-reg.nii.gz
Histogram calculation successful of 3020-t2w-xc-reg.nii.gz
Histogram calculation successful of 7015-t2w-xc-reg.nii.gz
Histogram calculation successful of 3005-t2w-xc-reg.nii.gz
Histogram calculation successful of 7027-t2w-xc-reg.nii.gz
Calculating histogram of 3011-t2w-xc-reg.nii.gz
Histogram calculation successful of 5002-t2w-xc-reg.nii.gz
Calculating histogram of 1026-t2w-xc-reg.nii.gz
Histogram calculation successful of 7001-t2w-xc-reg.nii.gz
Calculating histogram of 5009-t2w-xc-reg.nii.gz
Histogram calculation successful of 7022-t2w-xc-reg.nii.gz
Calculating histogram of 7014-t2w-xc-reg.nii.gz
Histogram calculation successful of 5011-t2w-xc-reg.nii.gz
Calculating histogram of 1021-t2w-xc-reg.nii.gz
Histogram calculation successful of 3009-t2w-xc-reg.nii.gz
Calculating histogram of 7028-t2w-xc-reg.nii.gz
Histogram calculation successful of 1023-t2w-xc-reg.nii.gz
Histogram calculation successful of 7006-t2w-xc-reg.nii.gz
Calculating histogram of 1027-t2w-xc-reg.nii.gz
Histogram calculation successful of 3018-t2w-xc-reg.nii.gz
Calculating histogram of 7008-t2w-xc-reg.nii.gz
Histogram calculation successful of 1026-t2w-xc-reg.nii.gz
Histogram calculation successful of 3011-t2w-xc-reg.nii.gz
Histogram calculation successful of 5009-t2w-xc-reg.nii.gz
Histogram calculation successful of 7014-t2w-xc-reg.nii.gz
Histogram calculation successful of 1021-t2w-xc-reg.nii.gz
Histogram calculation successful of 7028-t2w-xc-reg.nii.gz
Histogram calculation successful of 7008-t2w-xc-reg.nii.gz
Histogram calculation successful of 1027-t2w-xc-reg.nii.gz
Completed histogram calculation of all the subjects.
Time taken in histogram calculation 96.53516745567322 seconds
Time taken in feature extraction for training 96.73375797271729 seconds
```

In [95]:

```
# Test (predict goodness)
!python code/structuralQCbatch.py \
-i data/test_reg_list.txt \
-c data/test_cases.txt \
--fixedImg data/1001-t2w-xc.nrrd \
-e data/test_visual_qc.xlsx \
-o data/processed \
--type t2 \
--numThreads 16
```

```
antsApplyTransforms found
antsRegistrationSyNQuick.sh found
antsRegistration found
All executables are found, program will begin now ...
/home/tb571/ml_class/csci_e82_project/data/5010/raw/5010-t2w-xc-reg.nii.gz
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/1028/raw/1028-t2w-xc-reg.nii.gz
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/3024/raw/3024-t2w-xc-reg.nii.gz
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/5016/raw/5016-t2w-xc-reg.nii.gz
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/5015/raw/5015-t2w-xc-reg.nii.gz
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/7027/raw/7027-t2w-xc-reg.nii.gz
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/1017/raw/1017-t2w-xc-reg.nii.gz
/home/tb571/ml_class/csci_e82_project/data/5016/raw/5016-t2w-xc-reg.nii.gz
Registered image found ...
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/3025/raw/3025-t2w-xc-reg.nii.gz
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/3006/raw/3006-t2w-xc-reg.nii.gz
/home/tb571/ml_class/csci_e82_project/data/5012/raw/5012-t2w-xc-reg.nii.gz
Registered image found ...
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/3026/raw/3026-t2w-xc-reg.nii.gz
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/7018/raw/7018-t2w-xc-reg.nii.gz
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/1005/raw/1005-t2w-xc-reg.nii.gz
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/5003/raw/5003-t2w-xc-reg.nii.gz
/home/tb571/ml_class/csci_e82_project/data/3023/raw/3023-t2w-xc-reg.nii.gz
Registered image found ...
Registered image found ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
1 fail
/home/tb571/ml_class/csci_e82_project/data/3003/raw/3003-t2w-xc-reg.nii.gz
Registered image found ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/7026/raw/7026-t2w-xc-reg.nii.gz
Registered image found ...
2 fail
/home/tb571/ml_class/csci_e82_project/data/1031/raw/1031-t2w-xc-reg.nii.gz
Registered image found ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/7023/raw/7023-t2w-xc-reg.nii.gz
Registered image found ...
2 fail
/home/tb571/ml_class/csci_e82_project/data/3017/raw/3017-t2w-xc-reg.nii.gz
Registered image found ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/5024/raw/5024-t2w-xc-reg.nii.gz
Registered image found ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/7020/raw/7020-t2w-xc-reg.nii.gz
Registered image found ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/1006/raw/1006-t2w-xc-reg.nii.gz
Registered image found ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/3019/raw/3019-t2w-xc-reg.nii.gz
Registered image found ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/7023/raw/7023-t2w-xc-reg.nii.gz
Registered image found ...
Checking quality ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/7031/raw/7031-t2w-xc-reg.nii.gz
Registered image found ...
1 fail
/home/tb571/ml_class/csci_e82_project/data/1007/raw/1007-t2w-xc-reg.nii.gz
Registered image found ...
Checking quality ...
1 fail
/home/tb571/ml_class/csci_e82_project/data/5020/raw/5020-t2w-xc-reg.nii.gz
Registered image found ...
Checking quality ...
Checking quality ...
Checking quality ...
1 fail
/home/tb571/ml_class/csci_e82_project/data/5006/raw/5006-t2w-xc-reg.nii.gz
Registered image found ...
1 fail
/home/tb571/ml_class/csci_e82_project/data/7024/raw/7024-t2w-xc-reg.nii.gz
Registered image found ...
Checking quality ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/1018/raw/1018-t2w-xc-reg.nii.gz
Registered image found ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
Checking quality ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/5025/raw/5025-t2w-xc-reg.nii.gz
Registered image found ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/3012/raw/3012-t2w-xc-reg.nii.gz
Registered image found ...
4 pass
4 pass
/home/tb571/ml_class/csci_e82_project/data/7007/raw/7007-t2w-xc-reg.nii.gz
Registered image found ...
/home/tb571/ml_class/csci_e82_project/data/7028/raw/7028-t2w-xc-reg.nii.gz
Registered image found ...
4 pass
/home/tb571/ml_class/csci_e82_project/data/5013/raw/5013-t2w-xc-reg.nii.gz
Registered image found ...
2 fail
/home/tb571/ml_class/csci_e82_project/data/5027/raw/5027-t2w-xc-reg.nii.gz
Registered image found ...
Checking quality ...
1 fail
4 pass
Checking quality ...
Checking quality ...
4 pass
Checking quality ...
Checking quality ...
4 pass
Checking quality ...
2 fail
4 pass
4 pass
4 pass
4 pass
2 fail
2 fail
4 pass
4 pass
2 fail
4 pass
4 pass
Time taken in quality checking 460.45311093330383 seconds
```

In [96]:

```
%cat data/processed/confusion_matrix.txt
```

```
Confusion matrix (columns are true labels, rows are predicted labels):
Visual discrete scores are:  [1, 2, 3, 4]
[[ 5  0  0  0]
 [ 1  7  0  2]
 [ 0  0  0  8]
 [ 0  0  0 15]]
```

Observation

T2w image goodness check yielded as good as T1w image as characterized by the presence of less number of off-diagonal elements in the confusion matrix. Again, we can consolidate each quadrant of it as:

```
[[13  2]
 [0  23]]
```

From the bottom-corner entry, we conclude that we have been able to correctly predict the score of good images. On the other hand, 2 bad images has been misclassified as good ones.

Discussion

Similarity metric

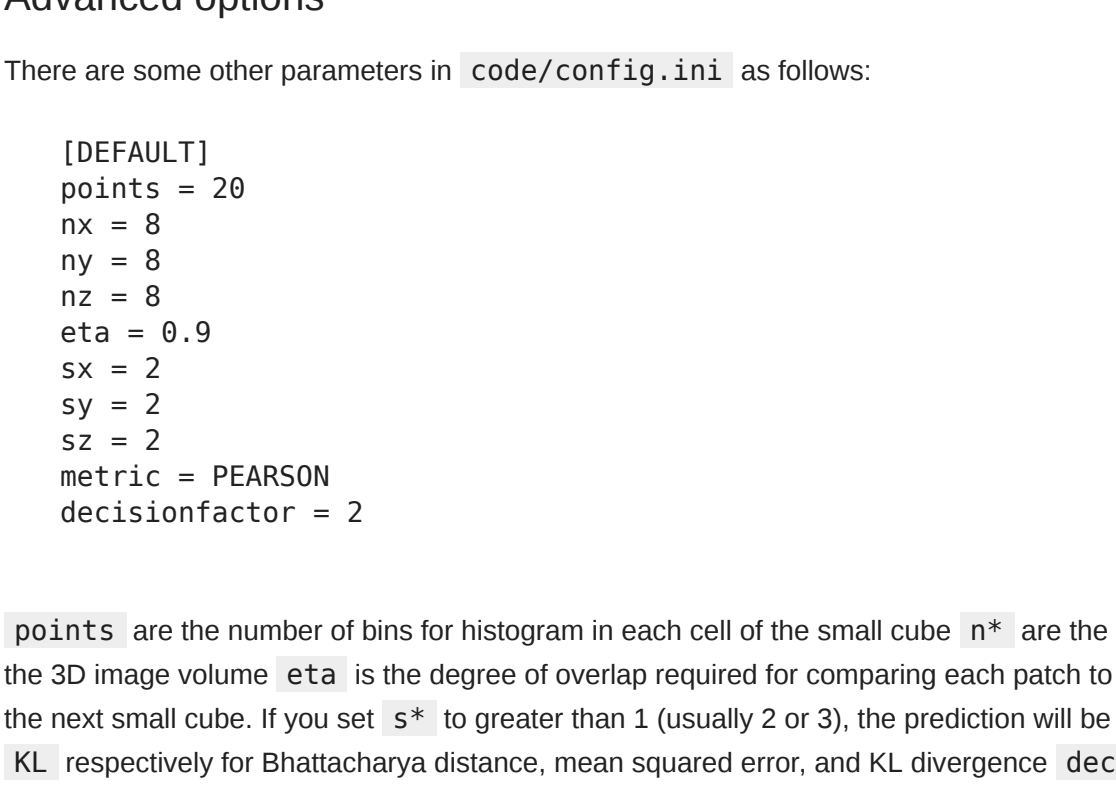
A remote user can try my algorithm by varying the similarity metric. The `metric = PEARSON` parameter in `code/config.ini` file has to be modified with one of `{'PEARSON','BC','KL','MSE'}`. I have tried all of them and submitted my best similarity metric for grading.

Cross Validation

The only cross validation I tried in this project is to find a proper kernel dimension `nx x ny x nz`. Although my algorithm allows varying numbers for `nx x ny x nz`, I have tried equal numbers i.e. cube. I have varied the cube dimension as `[3,5,8,10,20]` and decided on 8 from the following curve:

In [27]:

```
display(Image(filename="cv.png", height=400, width=600))
```



Majority voting

Since I have coded four similarity metrices:

```
* Pearson correlation coefficient
* Bhattacharyya distance
* Kullback–Leibler divergence
* Mean squared error
```

One would naturally think why I did not use a bagging/majority voting approach to decide a rating for the MRI. I tried that as well. However, bagging does not yield any better accuracy than a single similarity metric. It only complicates the code and probably confuses the algorithm. So I have omitted that untidy code from this submission.

Future work

Furnish bootstrapping

Currently, bootstrapping has been applied at the whole data level. Bootstrapping is synonymous to data duplication. So, there may be some samples that exist in both training and testing sets. Algorithmically, existence of a few train samples in the test set does not affect the outcome at all. But in future, we plan to write complex code that will apply bootstrapping to the train set only.

Use ROC AUC

In the above, we care more about pass/fail decision than the actual rating of the MRI (1,2 are fail while 3,4 are pass). Thus we have some buffer in evaluating performance of the algorithm. In such case, we can calculate probability of an MRI being a rating. Instead of hard assignment of a rating `[1,2,3,4]`, we can calculate probability as a four element array e.g. `[0.1,0.2,0.5,0.2]`. Such probabilities would allow AUC calculation from an ROC curve. We intend to use this metric in future.

Mask MRIs

A significant portion of the MRI is background i.e. black region around the brain. When comparing similarity between two MRIs, we should not need to consider that background. If we could extract the brain from an MRI, then the background would reduce to zero. The zero features should increase efficiency of calculation. In future, we shall look for a good brain extraction tool intrisically in Python that could be used conveniently in the above terminal.

Conclusion

Despite a few misclassifications, we have mostly been able to decide which MRI passes goodness check. Hence, this project has been a successful one. Upon completion of the future work, I plan to publish this at a machine learning conference.

Appendix

Advanced options

There are some other parameters in `code/config.ini` as follows:

```
[DEFAULT]
points = 20
nx = 8
ny = 8
nz = 8
eta = 0.9
sx = 2
sy = 2
sz = 2
metric = PEARSON
decisionfactor = 2
```

`points` are the number of bins for histogram in each cell of the small cube `n*` are the dimensions of the small cube that we slide along the 3D image volume `eta` is the degree of overlap required for comparing each patch to the library images `s*` is the stride taken to get the next small cube. If you set `s*` to greater than 1 (usually 2 or 3), the prediction will be faster `metric` currently supports `BC`, `MSE`, `KL` respectively for Bhattacharya distance, mean squared error, and KL divergence `decisionfactor` is the factor that determines threshold for predicting an image as pass. For example, if 1 is the worst and 4 is the best, then a `decisionfactor` of 2 will set halfway (2) as the threshold:

```
score_range= np.ptp(discreteScores)+1 if min(discreteScores)>0 else np.ptp(discreteScores)
quality= 'pass' if predicted_score>score_range/decisionFactor else 'fail'
```

Visual QC

In training mode, visual quality scores are required to train the algorithm. As occurs for most training, someone has to visually look at a batch of data and give it `discreteScores`. In our training, we have used `[1,2,3,4]` as the discrete scores while `1` is for worst i.e. an MRI affected by severe ringing, ghosting, motion, and signal drop. On the other hand, `4` is for a good quality MRI free from noticeable artifacts. The scores should be in an excel file as follows:

Subject ID	t1 score	t2 score
1001	1	4
1002	3	2
...		
...		

Make sure to properly write the header:

```
| Subject ID | t1 score | t2 score |
```

The caseids in the caselist should be same as the ones used under Subject ID column in the excel file.